Automatic music classification and the importance of instrument identification

Cory McKay, Dept. of Music Theory, McGill University, Canada
cory.mckay@mail.mcgill.ca

Ichiro Fujinaga, Dept. of Music Theory, McGill University, Canada
ich@music.mcgill.ca

Proceedings of the Conference on Interdisciplinary Musicology (CIM05)
Actes du Colloque interdisciplinaire de musicologie (CIM05)
Montréal (Québec) Canada, 10-12/03/2005

Abstract

This paper empirically demonstrates the particular effectiveness of features based on instrumentation in the realm of automatic music classification. The effectiveness of high-level features in general, relative to low-level signal-processing based features, is also demonstrated through an experiment involving automatic genre classification. Experimental evidence is also provided supporting the use of large feature sets combined with feature weighting systems.

The tools used to pursue this research are described, as is pertinent background information. An overview is provided of a large library of useful musical features that is available for research purposes to theorists, musicologists and other researchers. The Bodhidharma symbolic music classification system is also introduced as a useful and easy to use tool for researchers wishing to pursue research in music classification based on instrumentation and other high-level features.

Introduction to automated music classification

Automatic music classification is an area of research that has been receiving a great deal of attention in recent years. Music can be classified into taxonomies based on genre, performer, composer or geographical/cultural point of origin, to give just a few examples. Another approach is to cluster music based entirely on self-referential similarity, without reference to any pre-existing taxonomy.

Automatic music classification is a far from trivial task. Categorization based on genre or geographical region, for example, can be very difficult for both humans and computers, as there are very rarely precise, clear and consistent heuristics delineating the musical qualities and characteristic of each category. Indeed, it can often be difficult even to come up with a generally accepted taxonomy. The task of defining and implementing measurements of musical similarity can be equally murky. These ambiguities pose serious difficulties when researching music categorization, whether it be for theoretical or practical purposes.

Pattern recognition and machine learning techniques open up a wide range of opportunities for dealing with such issues. These techniques, for the most part, make use of statistical tools to search for and learn to recognize empirical regularities in collections of music. The results can then be combined with existing taxonomies, or used to construct new taxonomies, in order to assign meaning to these statistical regularities. Ultimately, this is used to form correlations between the characteristics of pieces of music and the categories of interest that can be generalized and used to classify new pieces of music.

Automated classification using machine learning has the practical advantage of allowing one to perform classifications much faster and cheaper than human manual classifications. It also has the theoretical advantages of analyzing music in original and non-intuitive ways that might not occur to humans while at the same time avoiding human biases and preconceptions that might contaminate experimental results.

There are a wide range of applications for automatic music classifiers, both scholarly and commercial. To give just a few examples, researchers could train classifiers to recognize the work of particular composers, and then use the trained classifiers to attempt to determine if compositions whose authorship is unknown are in fact likely to belong to be the work of a particular composer. Genre classifiers could be used to gain experimental insights into what characteristics are important in distinguishing between particular genres and how these characteristics vary. Classification results can be used to support sociological and psychological research into how humans construct the notion of musical similarity and form musical groupings, and how this compares to the “objective” truth produced by computerized classifiers. Classifiers of many varieties can be used to automatically analyze and sort recordings as they are added to large databases. Music recommendation systems can be used to sift through huge music databases of both well-known and obscure music and recommend unknown pieces to users based on what they are known to like. Classifiers can be trained to sort large personal music collections based on mood or on scenarios when one might like to listen to particular recordings, such as while driving or while doing the laundry. Similarity analysis can be used to help detect the distribution of pirated recordings. Automated transcription systems can be used to classify different sonic events as particular notes. These are only a few of the many possible applications of music classification systems.

Importance of features

One of the most important parts of automated classification is the choice of features, or properties, to extract from a piece of music. Features serve as the input to pattern recognition systems and are the basis upon which classifications are made. Even an optimal classifier cannot correctly classify music if it is not provided with features that properly segment recordings into the required taxonomy.

Features can be extracted from audio recordings, such as MP3’s or CD audio files, or from symbolic recordings in formats such as MIDI or Humdrum. Correspondingly, features may be low-level, which is to say that they are based on signal processing quantities that have little intuitive meaning to humans, or they may be high-level, which is to say that they make use of musical abstractions that are meaningful to humans. Examples of high-level features include tempo and meter. Examples of low-level features include zero-crossings and spectral frequency ratios. It is currently difficult or impossible to reliably extract many high-level features from audio recordings, therefore making it necessary to use symbolic recordings if one wishes to reliably obtain a variety of high-level features.

The critical role that features play in automatic classification makes them worthy of particular attention. Knowledge of which features work best in particular contexts can significantly improve classification results. Furthermore, an analysis of which features are important in making particular classifications can be an important source of theoretical musicological information in and of itself. This is the motivation for the research described in this paper.

Overview of this research

This research is concerned with studying the effectiveness of different high-level features when applied to the task of genre classification. High-level features in particular were chosen as the subject for this research because they have direct musical meaning that is much more useful and meaningful to musicologists and theorists than signal processing statistics. This is not to say that low-level features are not useful both practically and from a psychoacoustic research perspective, but it is very difficult to deduce theoretically meaningful musical information from them.

The particular task of genre classification was chosen because it is a difficult task that forces classifiers to deal with a significant amount of ambiguity and inconsistency. Although it is certainly true that there is no guarantee that features that are useful for one task will be useful for another task, there is no way to exhaustively test all classification problems. Many classification problems are indeed sub-sets of the genre classification problem, and the difficulty of the genre problem implies that successful features in this context are likely to be particularly robust, with the implication that they have an increased probability of being useful for other types of classification.

As will be described in more detail in the following sections, the experiments performed here used genetic algorithms to empirically weigh various features wile training genre classifiers. It was found that features relating to instrumentation performed significantly better than any other class of features. This is pertinent because instrumentation is in essence a high-level abstraction of certain timbral characteristics. It is important to consider ways in which we conceptualize timbre, and instrumentation is an important example of such a conceptualization.

Related research

There has been a great deal of material published on machine learning, pattern recognition and classification in general. The books of Alpaydin (2004), Duda, Hart and Stork (2001) and Russell and Norvig (2002) provide particularly good complementary overviews. Sakawa (2002) provides good background information on genetic algorithms.

Although there has certainly been a great deal of theoretical work published on useful musical characteristics for specific types of music, empirical research on features that are useful for arbitrary types of music has been largely neglected. Fortunately, there have been a few exceptions. The Cantometrics project (Lomax 1968) made an important contribution towards understanding the relevance of various features across cultures. Arden and Huron (2001) studied the correlation between different musical features and different geographical regions. Cope (1991) has published some useful features, and Tagg (1982) has published a good theoretical overview of techniques that can be applied to a variety of musics.

An understanding of the many issues relating to how different cultures and sub-groups define musical categories, construct taxonomies and form interrelationships between categories is essential to approaching classification research. These issues have been well discussed in the work of Fabbri (1982, 1999), Brackett (1995), Frith (1996), Pachet and Cazaly (2000), Aucouturier and Pachet (2003) and McKay (2004). Lakoff (1987) also provides a very good resource on issues relating to categorization and labelling in general. Although a detailed discussion is beyond the scope of this paper, these issues are important to consider when dealing with automatic music classification, as a taxonomy that is not meaningful and useful to a variety of humans in a variety of scenarios is of little general value. One must also be careful to consider how taxonomies change with time, in terms of the categories used, their memberships and their interrelationships.

Most of the research to date in automatic music classification has concentrated on audio recordings. Perhaps the best-known of this work is by Tzanetakis and Cook (2002). Classification of symbolic recordings has been studied to a much lesser extent, with the best results to date having been achieved by McKay and Fujinaga (2004). McKay (2004) provides a more complete survey of both audio and symbolic music classification systems. To the best of the authors' knowledge, no experimental evidence has been published to date regarding the relative effectiveness of a large body of high-level features when applied to automatic music classification tasks.

Methodology for evaluating feature effectiveness

The research carried out here was performed using an updated version of the Bodhidharma MIDI classification system (McKay 2004). This system utilizes sophisticated ensembles of neural networks and k-nearest-neighbour classifiers to classify symbolic recordings based on arbitrary taxonomies. This system has been shown to produce very good results (McKay and Fujinaga 2004).

Bodhidharma is a free, flexible and easy to use software tool that researchers in musicology, music theory and music technology can use to study arbitrary types of music classification and the effectiveness of various features in a variety of contexts. Users can easily enter their own taxonomies and recordings using the GUI, and it is a relatively easy matter for more technically proficient users to add their own features. Figure 1 shows a sample screen shot of the Bodhidharma interface.

The Bodhidharma system was particularly useful for this research because it utilizes a large collection of high-level features and because it performs feature weightings on all features while training. Feature weighting is a technique that can be used to aid the performance of classifiers. Many classifiers operate as "black boxes" that take in features and produce classifications, without revealing the precise mechanisms involved. Providing a large number of features to a classifier has the advantage of giving it a large number of percepts on which classifications can be based, but has the disadvantage of potentially overwhelming the classifier with so much information that it is unable to effectively recognize patterns. Feature weighting is a means of attempting to find a balance between these concerns, by experimentally weighting the importance a classifier pays to each feature, and potentially eliminating the features that appear least likely to be useful.

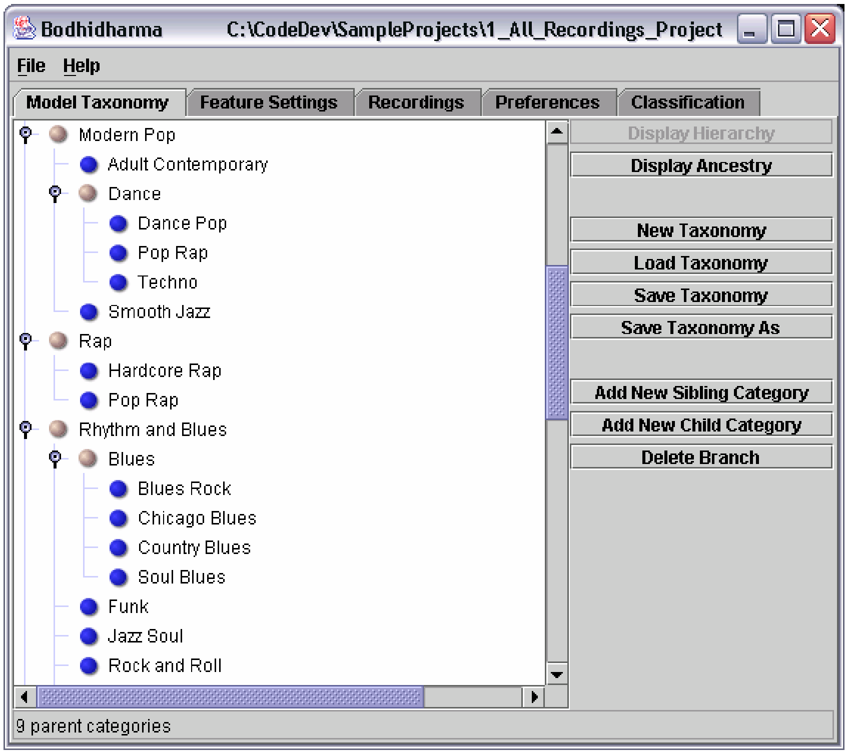

Figure 1. A screenshot of part of the Bodhidharma MIDI classification interface.

Bodhidharma uses genetic algorithms (GA's) to arrive at feature weightings. GA's operate by loosely modelling biological evolution. Pieces of data (feature weightings, in this case) are encoded as "chromosomes." A number of chromosomes are randomly generated, and the "fitness" of each chromosome is evaluated based on, in this case, the classification effectiveness of a particular set of feature weightings. Some chromosomes are selected to "breed," with a likeliness to be selected based on their fitness. The breeding chromosomes exchange pieces of their data with each other, with random "mutations" sometimes introduced, in order to produce a new generation of chromosomes. This process continues iteratively, with fitnesses generally increasing from generation to generation, until a chromosome is produced that represents an acceptable solution.

GA's essentially search the set of possible feature weightings in order to produce a good, although not necessarily optimal, solution. Although one would ideally like to perform exhaustive searches of all possible feature weightings to arrive at optimal solutions, this problem is computationally intractable for large feature sets, and could well result in overtraining that limits generalization to recordings outside the training set. GA's are an effective way of finding good solutions to complex optimization problems like feature weighting.

The application of GA's to feature vectors can be interpreted as an experimental evaluation of the effectiveness of different features, as represented by their weightings, for the classification task at hand. These weightings, as averaged over many trial experiments, were used as the experimental results in this research project.

It should be noted that Bodhidharma performs hierarchical and round robin classification as well as flat classifications. This means that separate feature weightings were found when classifying among all categories, as well as when classifying among all pairs of categories individually and among only the descendants of particular parent categories in a hierarchical taxonomical structure. This is a very useful aspect of Bodhidharma for musicological and theoretical research, as it allows one to find not only the best features overall, but the relative effectiveness of different features when classifying among different category sub-sets.

Features used

The importance of features justifies that they be given particular attention. One would like to have a wide range of features available in order to provide classifiers with as much useful information as possible, as well as to provide researchers with different bases that they can use to analyze music. At the same time, however, it is usually desirable to avoid redundant or irrelevant information that could complicate the tasks of both classifiers and researchers.

In the context of general-purpose analysis and classification, one wishes to have features that are applicable to a wide range of musics and that do not make theoretical assumptions that limit their scope. Simple and descriptive measurements of overall characteristics and variances in a piece of music can be particularly effective in this respect, and tend to be more general than features derived from sophisticated theoretical analyses. Considerations from a variety of fields are relevant to the choice of features but, as can be seen from the relatively limited number of references on general musical features in the Background Research section of this paper, relatively little work has been published on specific features that can be usefully extracted from arbitrary types of music.

The Bodhidharma system comes packaged with 111 features that can be extracted from MIDI recordings. These features have been collectively designed to be effectively applicable to as wide a range of musics as possible. These features fall into the following six groups:

- **Instrumentation:** What types of instruments are present and which are given particular importance relative to others? The importance of non-pitched instruments and their interaction with pitched instruments is considered.
- **Texture:** How many independent voices are there and how do they interact (e.g. polyphonic, homophonic, etc.)? What is the relative importance of different voices?
- **Rhythm:** The time intervals between the attacks of different notes and the durations of each note are considered. What kinds of meters and rhythmic patterns are present? Is rubato used? How does rhythm vary from voice to voice?
- **Dynamics:** How loud are notes and how much variation in dynamics is there?
- **Pitch Statistics:** What are the occurrence rates of different notes, in terms of both MIDI pitches and pitch classes? How tonal is the piece? What is the range? How much variety in pitch is there?
- **Melody:** What kinds of melodic intervals are present? Is there a lot of melodic variation? What kinds of melodic contours are used? What types of phrases are used and how often are they repeated?

It should be noted that two types of features are implemented in Bodhidharma: one-dimensional and multi-dimensional features. One-dimensional features each consist of a single number that represents an aspect of a recording in isolation. Multi-dimensional features consist of sets of related values that have limited significance taken alone, but together may reveal meaningful patterns. This separation was used both to accommodate the characteristics of the different kinds of information being considered and because the use of one-dimensional features with k-NN classifiers and multi-dimensional classifiers with neural nets proved to be fruitful based on the characteristics of these classifiers.

This large library of features provides a valuable resource for researchers who wish to analyze or classify arbitrary types of music for whatever reasons they wish. It is particularly valuable in that it was designed with the particular goal of not making any theoretical assumptions that would limit its scope to particular types of music. In addition, another 42 features have been proposed but not implemented in Bodhidharma. Although too numerous to describe individually here, precise descriptions of each feature as well as the design philosophy behind the feature library as a whole are available elsewhere (McKay 2004).

Experimental details

Two hierarchical taxonomies were used to train the classifiers and evolve feature weightings. One consisted of 9-leaf categories and the other consisted of 38-leaf categories. The first taxonomy was designed to provide a set of categories comparable in size to taxonomies used in previous automatic genre classification research, and the second was designed in order to allow tests under more realistic conditions. These taxonomies are shown in Figures 2 and 3.

Although care was taken when designing these taxonomies to cover a wide variety of genres using a realistic structure and set of categories, these taxonomies are intended only for test purposes. The structure used deviates from standard hierarchical structures in that categories may belong to more than one parent and recordings can belong to more than one category. Although this significantly complicated the classification task, it was necessary in order to simulate realistic genre scenarios.

Jazz
- Bebop
- Jazz Soul
- Swing

Popular
- Rap
- Punk
- Country

Western Classical
- Baroque
- Modern Classical
- Romantic

Figure 2. Reduced taxonomy.

Country
- Bluegrass
- Contemporary
- Traditional Country

Jazz
- *Bop*
  - Bebop
  - Cool
- *Fusion*
  - Bossa Nova
  - Jazz Soul
  - Smooth Jazz
- Ragtime
- Swing

Modern Pop
- Adult Contemporary
- *Dance*
  - Dance Pop
  - Pop Rap
  - Techno
- Smooth Jazz

Rap
- Hardcore Rap
- Pop Rap

Rhythm and Blues
- *Blues*
  - Blues Rock
  - Chicago Blues
  - Country Blues
  - Soul Blues
- Funk
- Jazz Soul
- Rock & Roll
- Soul

Rock
- *Classic Rock*
  - Blues Rock
  - Hard Rock
  - Psychedelic
- *Modern Rock*
  - Alternative Rock
  - Hard Rock
  - Metal
  - Punk

Western Classical
- Baroque
- Classical
- *Early Music*
  - Medieval
  - Renaissance
- Modern Classical
- Romantic

Western Folk
- Bluegrass
- Celtic
- Country Blues
- Flamenco

Worldbeat
- *Latin*
  - Bossa Nova
  - Salsa
  - Tango
- Reggae

Figure 3. Full taxonomy.

A total of 950 MIDI files were collected and hand-classified for use in training and testing. These recordings were selected in order to include a variety of sub-types within each genre category. The Baroque category, for example, included operas, cello suites, violin concertos, organ chorales, etc. Although this made the task of classification much more difficult, it was necessary in order to ensure that realistic conditions were simulated. Model classifications for training and testing purposes were made primarily based on a combination of the classifications made on a variety of on-line sites.

Training and testing was performed using five-fold cross-validation. This means that all results presented in the following section are the averages of five different experiments, each involving different partition of the MIDI files between training and for testing.

Results and discussion

The genre classification experiments resulted in very good success rates. The system was able to correctly classify MIDI recordings among 9 categories 90% of the time and among 38 categories 57% of the time. Furthermore, the root genre (boldfaced in Figures 2 and 3) was correctly identified 98% of the time for the 9 taxonomy and 81% of the time for the 38-category taxonomy. These results are significantly better than those produced by any other symbolic genre classification system published to date, and at least as good as existing audio classification systems, which have generally achieved success rates below 80% when classifying among more than five categories. In addition, the 38-

category problem involved a much larger and more sophisticated taxonomy than any other system has dealt with to date. The results for the 9-category taxonomy are shown in more detail in Figure 4.

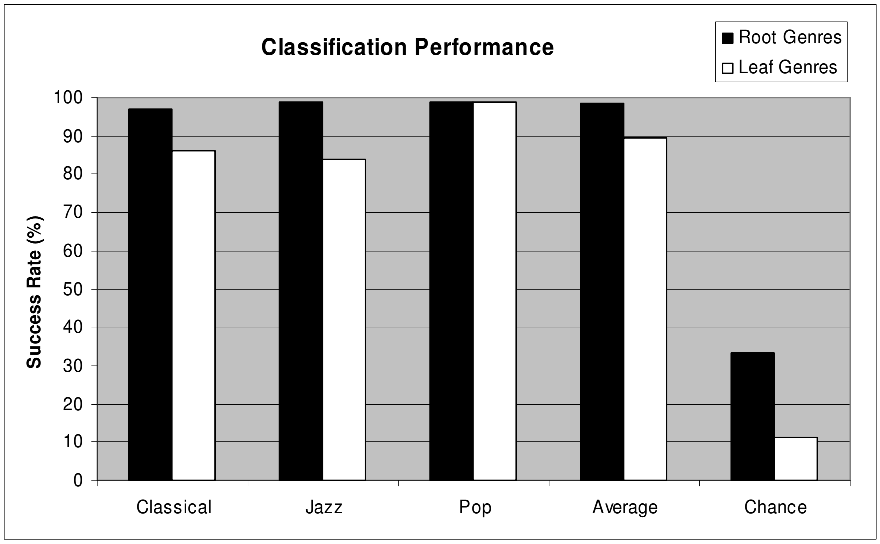

Figure 4. Average classification success rates. The leaf genre bars give the average success rates of the leaf genres belonging to the corresponding root genres.

These results demonstrate not only the effectiveness of Bodhidharma and its classification system, but also that high-level features can produce very good results, and should not be neglected in favour of low-level features, as has largely been the case in the automatic music classification literature. These results also show that very good results can be achieved with small taxonomies, and that we are at least approaching a point where large taxonomies can be dealt with effectively.

The results regarding the relative importance of different features and feature groups were very interesting. The top seven features together were assigned a total, on average, of 92% of the weightings assigned to the 111 features used. These seven features are described in Table 1.

One might be tempted, upon viewing Table 1 and considering the dominance of these top seven features, to conclude that one need not bother with the additional 104 features. This is not at all the case, however. To begin with, although the experiments performed here involved a large number of categories and many trials, and the results are likely generalizable to a certain extent, there is no guarantee that any particular feature that worked well here would always be a top performer for any arbitrary taxonomy.

Furthermore, an additional experiment was performed where a varying number of randomly chosen features were removed from consideration by the classification system. The results of this experiment are shown in Figure 5. It is clear from this figure that increasing the number of candidate features significantly improves performance. Of course, it could be argued that this is because the top performing features as outlined in Table 1 were more likely to be absent in the smaller feature sets, and it is their absence that caused the drop in performance. This is not, however, the case. Upon detailed examination of the feature weights assigned under different trials and different classifiers, it was found that there were indeed a number of cases where the seven classifiers described in Table 1 played only a minor role.

In order to understand this, one must recall that the classification system used involved a combination of both hierarchical and round-robin classification as well as basic flat classification, and that classification at each of these levels involved separately evolved feature weightings. It is not at all surprising, for example, that entirely different features would be useful for distinguishing between Baroque and Classical music and between Chicago Blues and Country Blues.

Group	Feature Name	Weight	Feature Description
I	Time Prevalence of Pitched Instruments	24.1%	A features array with one entry for each of the 128 General MIDI Instruments. Each entry was set to the total time in seconds during which a given instrument was sounding divided by the total length in seconds of the piece.
P	Pitch Class Distribution	17.5%	A features array with a bin for each pitch class. Magnitudes of each bin corresponded to the normalized number of notes occurring in the piece belonging to the corresponding pitch class.
I	Unpitched Instruments Present	16.7%	A features array with one entry for each of the 47 MIDI Percussion Key Map instruments. Each entry was set to 1 if at least one note was played using that patch and to 0 if the patch was not used.
R	Initial Time Signature	8.8%	A features array with two elements. The first was the numerator of the first occurring time signature and the second was the denominator of the first occurring time signature.
P	Basic Pitch Histogram	8.5%	A features array with one bin for each MIDI pitch, with each bin having a magnitude corresponding to the normalized number of notes in the piece at the corresponding pitch.
R	Beat Histogram	8.4%	A features array with entries corresponding to the relative frequency of different note onset autocorrelations.
M	Melodic Interval Histogram	8.2%	A features array with bins corresponding to the number of melodic semitones between adjacent notes on a given MIDI channel. Bin magnitudes indicated the normalized number of times note onsets were separated from the previous note by the given interval.

Table 1. The seven best performing features, on average. The Weight column indicates the average percentage of total weightings (among all 111 features) assigned to each feature.

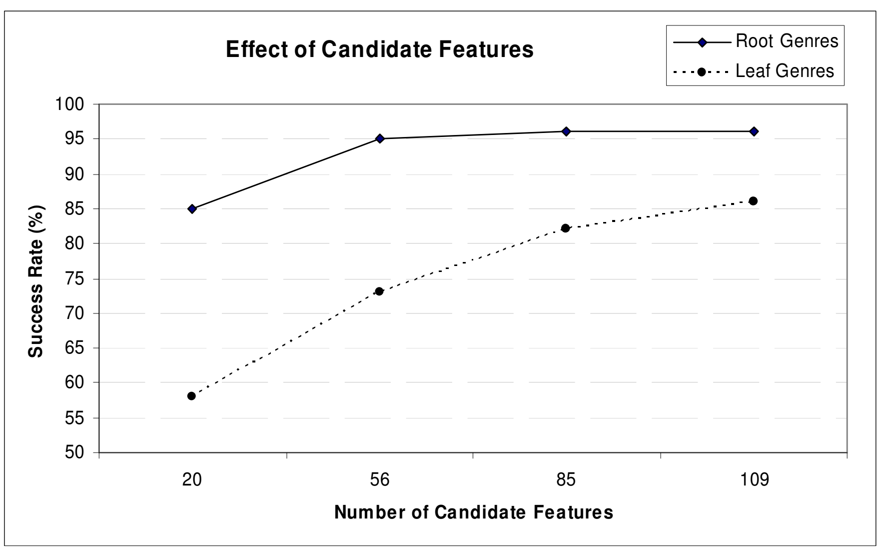

Figure 5. Effect, on average, of limiting the number of features provided to classifiers.

Considering the wide variety of the 38 categories, it is more than reasonable that there were some important deviations in useful features from category to category. So, while the seven features in Table 1 were certainly dominant overall, and are worthy of particular attention, features that were assigned a relatively insignificant weighting overall played a significant or even dominant role in particular types of classification. This means that the removal of features with low overall weightings could have a serious negative impact on classification success rates.

One can observe from Table 1 that two of the top three performers overall were based on instrumentation. Furthermore, as can be seen from Table 2, instrumentation features overall were assigned 41.8% of all raw weightings and, when scaled by the proportion of features in each group, 46.1% of the weightings.

Feature Group	Combined Weighting (%)	Number of Features	Weighting Scaled by Feature Count (%)
Instrumentation	41.8	20	46.1
Texture	1.1	14	1.7
Rhythm	19.5	30	14.3
Dynamics	0.3	4	1.6
Pitch	27.8	25	24.5
Melody	9.5	18	11.6

Table 2. Combined weightings assigned to different feature groups. The Combined Weighting column indicates the combined weightings assigned to all features in a group, on average. Number of Features indicates the total number of features belonging to each group. The Weighting Scaled by Feature Count column indicates the combined weighting of all features in each group after scaling by the fraction of total features in each group.

It is therefore apparent that features based on instrumentation are by far the most significant group of features, with the closest competitor, pitch-based features, being assigned barely half the weightings after scaling. Furthermore, detailed examination revealed that at least one instrumentation feature played a major role in almost every classifier. So, while this certainly doesn't mean that other types of features should be neglected, it does indicate the particular importance of instrumentation in automatic music classification.

Conclusion

The research presented in this paper demonstrated empirically that high-level features can be very effective in classifying music. This is an important result, as much automatic music classification research to date has focused on low-level features. A genre classification experiment using only high-level features successfully classified MIDI recordings among 9 categories 90% of the time and among 38 categories 57% of the time, success rates that compare favourably to previous research with both symbolic and audio data.

This paper also demonstrated that a large library of features can significantly improve performance over smaller libraries when coupled with feature weighting techniques and hierarchical and/or round robin classification.

A variety of features were analyzed in terms of relative overall effectiveness, and it was found that a few (7 out of 111) played a dominant role, but that even relatively minor features overall played essential roles in particular types of classification.

Of particular interest, it was found that features based on instrumentation (i.e., timbre) were by far the most important type of features. They were collectively assigned 41.8% of the importance among the six feature groups, a value that increases to 46.1% when scaled based on the number of features in each group. Not only were instrumentation-based features assigned almost twice the weighting collectively of the second-best performing feature group, but two of the top three individual features were based on instrumentation.

The importance of instrumentation implies the importance of timbre in general, as the human perception of instrumentation when listening to audio signals is essentially an abstraction extracted from timbral information. It is reasonable to suggest that future research and analysis, both theoretical and practical, should pay particular attention to instrumentation. This is a very interesting result, given that instrumentation has often been relegated to a secondary role in the literature.

Furthermore, these results imply that it might well be useful to prioritize the development of instrument identification sub-systems for audio classification systems, as the resultant high-level instrumentation features could prove more profitable than the low-level timbral features that are currently used in audio classification.

Acknowledgements

The generous financial support of the Social Sciences and Humanities Research Council of Canada has helped to make this research possible.

References

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

Alpaydin, E. 2004. *Introduction to machine learning*. Cambridge, MA: MIT Press.

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

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

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

Duda, R., P. Hart, and D. Stork. 2001. *Pattern classification*. New York: John Wiley & Sons Inc.

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

Fabbri, F. 1999. Browsing music spaces: Categories and the musical mind. *Proceedings of the International Association for the Study of Popular Music Conference*.

Frith, S. 1996. *Performing rites: On the value of popular music*. Cambridge, MA: Harvard University Press.

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

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

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.

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

Russell, S., and P. Norvig. 2002. *Artificial intelligence: A modern approach*. Upper Saddle River, NJ: Prentice Hall.

Sakawa, M. 2002. *Genetic algorithms and fuzzy multiobjective optimization.* Norwell, MA: Kluwer Academic Publishers.

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

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