Recognition of Musically Similar Polyphonic Music

Michael Chan and John Potter
School of Computer Science & Engineering
University of New South Wales
Sydney, NSW, 2052, Australia
{mchan, potter}@cse.unsw.edu.au

Abstract

When are two pieces of music similar? Others have tackled this problem either by considering the acoustic signals of musical performances, or by looking at features of a symbolic rendition of the piece, either as MIDI data or as some direct representation of the music score. This paper presents a new approach to assessing the similarity of polymorphic music segments by combining a feature-driven clustering approach with one that measures the contrapuntal similarity of the segments. On a composer classification task, our techniques achieved almost 80% accuracy when applied to a large database of short music segments from four classical composers. This is a significant improvement to other work on composer classification based on melodic themes.

1. Introduction

The task of recognizing when two pieces of music are similar is challenging, made more complicated by polyphony. Polyphonic music is composed of two or more parts, for example, a melody with accompanying base chords, or a four-part chorale or fugue. In this paper, we address the problem of deciding when two segments of polyphonic music can be considered musically similar. Specifically, we focus on the notion of similar *musical intuition*, which is motivated by [7] and defined to be a form of *unconscious knowledge that the listener brings to his hearing*. We do not consider surface patterns as such, however. We reduce the complexity of the task by restricting the length of the segments to just a few bars; we improve the discrimination of the musical pattern recognition, by exploiting the contrapuntal structure of the polyphonic music.

There are many potential applications for being able to recognize the similarity between two pieces of music. These include the automation of tasks such as music-genre classification and composer identification, and the development of computer-assisted music appreciation, improvisation and composition aids. In fact, the work presented in this paper has formed part of a larger effort building towards an automated system for composing music in a style sensitive manner, which we hope to report on elsewhere.

In Section 2, we describe the components of our approach, which is based on a combination of rhythm and pitch features of the segments, together with a technique for comparing the contrapuntal structure of the segments. We use an automatic classification approach to infer the composer for the segment, which can be further used to discriminate between segments. In Section 3 we indicate how the components are combined to produce our recognizer.

The question of whether some segments of music sound similar to others is a cognitive question, and is necessarily a subjective one. As part of the evaluation of our technique, we have therefore undertaken experiments with human subjects, to see if they agreed with the similarity groupings produced using our technique. A more objective measure of the effectiveness of our approach comes with composer identification. The results of our evaluations are described in Section 4.

Many researchers have studied musical similarity measures by analyzing acoustic signals. For example, Aucouturier et al. [1] introduced a musical timbre similarity measure based on Gaussian Mixture models of mel-frequency cepstrum coefficients (MFCCs). In addition, Foote in [5] describes a music indexing system also based on MFCCs, but instead relies on a supervised vector quantizer.

Much work has also been devoted to analyzing MIDI data. For example, Ghias et al. [6] introduced a system that searches a MIDI database by humming into a microphone. The hummed query is defined by a three-step representation and matched against the records in the database; technically, the melodic contour used should be regarded as the pitch contour, at least it is not of the (more genuine) type described by Temperley in *The Cognition of Basic Musical Structures* (CBMS) [13], which models *melodic streams*. Music style classification based on MIDI recordings is also

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

widely explored, e.g. McKay [9]. McKay devised a considerably large set of musical features describing music that is in MIDI format and applied them to classify music genre; our technique relies on some of these features.

2. Modeling Musical Attributes and Music Cognition

Unfortunately, modeling music is vastly different from many other modeling problems because music is obviously highly subjective. We must be careful when choosing the attributes to be modeled because those are the only sources from which knowledge can be learnt. For example, unlike weather, which can be modeled based on attributes such as temperature, humidity, air pressure, and so forth; music also has such quantitative parameters, such as rhythm and pitch, but unconscious musical perception and cognition play significant roles in determining quality of music. So, in order to model music cognition, we adopt a music cognition and psychology oriented structure used in the CBMS model.

2.1. Musical Features

Features are particularly useful in describing quantitative attributes. As mentioned, McKay devised a large features library containing 160 features for his music genre classification system, and many of the features used for our technique come from this library. These features are categorized into *instrumentation*, *musical texture*, *rhythm*, *dynamics*, *pitch statistics*, *melody*, and *chord*.

Although McKay's feature library is extensive, not all features are in fact useful for our purpose. In fact, more features do not necessarily lead to better performance because of the phenomenon known as the *curse of dimensionality* [4]. When new features do not provide useful information, data points can easily be separated and made sparser, and hence become harder to recognize. One way to overcome the curse of dimensionality is to reduce the feature set (thus, dimensionality) by incorporating prior knowledge in order to select useful features. We have selected a relatively small subset of features based on our subjective assessment of their potential usefulness.

After considering the practicality and plausibility of each feature, we have chosen to restrict attention to the *rhythm* and *pitch statistics* categories. Clearly, rhythm and pitch are fundamental and crucial elements of music. This is not to say that other categories are not important musical attributes, but they seem relatively less important for our task. For example, composers are usually not confined to producing music for some particular instruments or music of some particular textures, and therefore information on *instrumentation* and *musical texture* is not likely to be useful in composer identification.

Table 1. Features selected for our method

Type	Feature
Rhythm	`R-15` – `R-23`
	Avg. note duration above mid. C
	Var. of note duration above mid. C
	Avg. note duration below mid. C
	Var. of note duration below mid. C
	Avg. time between attacks above mid. C
	Var. of time between attacks above mid. C
	Avg. time between attacks below mid. C
	Var. of time between attacks below mid. C
Pitch stats.	`P-1` – `P-6`, `P-8` – `P-12`

Figure 1. Measures 14-16 of Chopin's Mazurka no.50 in A min., B.134.

Besides adopting features from McKay's library, we have introduced eight other features. All features used are summarized in Table 1, in which features prefixed with `R` or `P` come from McKay's feature library.

2.2. Contrapuntal Structure

The contrapuntal analysis in CBMS is a method to devise a structure outlining the melodic lines in a music passage. Based on the theory of stream segregation, devised by Bregman [2], the contrapuntal structure targets the subprocess of sequential integration, which Bregman defines as *putting together events that follow one another in time*. Contrapuntal structure is concerned with music perception and cognition of music listeners, and so conveys more information about the cognitive underpinnings of the music than do simple structures based merely on pitch or rhythm. Figure 2 depicts a contrapuntal structure of the excerpt of music shown in Figure 1.

The flow of melodic streams of music segments should reveal interesting perceptive information, which can be heard by listeners. Therefore, if melodic streams can characterize music segments, comparison of the contrapuntal structure of two music segments should determine whether they exhibit similar melodic structures, and also whether they would be perceived to be similar. Other analyses, including [3] and [10], have simply focused on pitch structures. Our novel technique models melodic streams using a more cognition based approach, and therefore should cap-

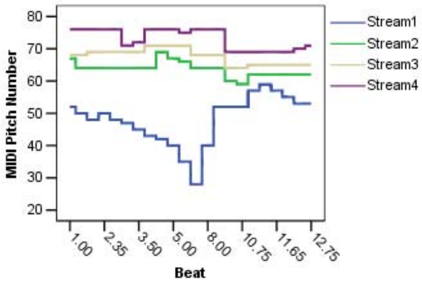

Figure 2. Contrapuntal structure of Figure 1.

ture musical intention more accurately. Although CBMS is designed for analysing Western classical music, it can be generalised to other styles of music such as Rock and African music.

In order to compare between contrapuntal structures, we compute the Levenstein distance [8] between them. Levenstein distance is commonly used for inexact matching between two strings and is particularly useful in spell checkers. Our matching technique is based on Huron's idea of measuring the similarity between two pitch structures [10], whereby the Levenstein distance between them is calculated. The technique is not vastly different from the original technique for two strings because melodic streams can be modeled using two dimensions, one representing pitch number and one representing onset-time (Figure 2).

2.3. Composer Classification

We also incorporate the knowledge of the classified composer according to a classifier in order to increase the descriptive power of our model. Using the features described, a Naïve Bayes (NB) classifier is trained to classify the composer of a given music segment. At the music segment level, elements of composer's style can be relatively subtle compared to those at a higher level, such as the phrase and the section levels. Given a small music segment, it can be difficult for even an experienced music listener to make such a classification because such segment is likely to lack clear structure, harmonic progressions, and so forth. Nevertheless, it is feasible for a segment of a Chopin piece to be musically similar, or simply have similar features, to a Bach segment, and hence the knowledge of composer classification on music segments may shed some new light on automatic music analysis.

3. Recognizing Similar Musical Intuitions

The crux of our technique for recognition of segments having similar musical intuitions lies in performing a clustering analysis on the three attributes described. Given the sort of model we have, one appropriate clustering technique is GDBSCAN (Generalized DBSCAN) [12], because not only can it produce arbitrarily shaped clusters, it also allows non-spatial constraints. Non-spatial constraints are useful here because both contrapuntal structure comparison and composer classification are non-spatial attributes; musical features are spatial attributes and are standardized.

Two music segments only belong to the same cluster if all three constraints can be satisfied. The first constraint governs the degree of similarity of the spatial attributes (musical features) by checking whether the Euclidean distance between them is within a certain threshold. The second and third constraints focus on the non-spatial attributes by respectively checking if the Levenstein distance between the contrapuntal structures is lesser than a certain amount and if both segments result in the same composer classification. Hence, segments having similar musical intuitions can be quantitatively defined to be those segments that have similar musical features and melodic streams and are in the same composer classification. Since the degree of similarity in musical features and melodic streams depend upon predefined numeric parameters, our technique therefore supports both near-exact (if the parameters are equal to zero) and inexact matching.

4. Evaluation

Experiments were conducted to evaluate both the performance of classifying the composer of segments using a classifier trained by the selected musical features and the precision of recognizing segments with similar musical intuitions. The former was evaluated by an objective assessment, whereas the latter required a subjective assessment because it is best for human listeners to identify musical similarity/dissimilarity.

4.1. Identifying Composers of Music Segments

With this experiment, we aimed to investigate the discriminating power of the features, in terms of composer classification. To this end, we trained and tested four classifiers: 1-nearest neighbor (1-NN), 3-nearest neighbor (3-NN), C4.5, and NB. Each was run with 5-fold validation on a collection of 28,225 music segments, which comprised 6,102 Bach, 10,096 Beethoven, 6,049 Chopin, and 5,978 Brahm segments of two to six beats long. Table 2 shows the success rates after five runs.

Although 1-NN yielded the best result, a NB classifier was actually chosen for checking the composer classification constraint in our implementation simply because the NB classifier turned out to be more efficient. The classifiers

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

Table 2. Rates of successful classification

Classifier	Success %
1-NN	79.4
C4.5	75.3
NB	74.9
3-NN	71.2

Table 3. Statistics of evaluation results.

N	Min.	Max.	Mean	SD
5	79.3%	88.6%	84.4%	3.7%

performed in the success rate range of 71%-79%, which is significantly higher than the average result (42%) produced by Pollastri et al. [11], based on 605 themes from five composers, from Mozart to the Beatles. Their work is nevertheless one of the first research efforts in automatic classification of composer. Although it aims at classifying from melodies, which carry different information, it is the closest comparison for our results we have been able to find.

4.2. Discriminating Segments in a Mixed Collection

In order to estimate the quality of our overall model, we invited five human subjects with at least three years of musical training to listen to groups of segments that are identified as similar by the technique described. Subjects were not given specific criteria for their judgments, but were given the definition of 'musical intuition'; this experiment is necessarily highly subjective. Each subject listened to 16 recordings, each containing four random unique segments that are identified as similar by our system, using the same dataset as the previous experiment. After playing each recording, subjects were required to note down whether they perceived all of the four segments to be similar. Table 3 shows the mean percentage and other statistics of segments identified as similar and agreed by subjects.

Despite the preliminary nature of this experiment, the results for comparing between the judgments made by the subjects with those made by the system are encouraging and confirm that our proposed technique is, at least, a promising candidate as an approach to recognizing the similarity of polyphonic music.

5. Conclusion and Future Work

Described is a novel technique to identify segments of MIDI recordings that convey similar musical intuitions. As part of the technique, a classifier is trained to classify composers from segments of music. We evaluated the classification process with a dataset containing more than 28,000 segments and achieved a success rate of 74.9% using a Naïve Bayes classifier. For the overall technique, a subjective assessment based on the same dataset showed that, on average, human listeners concurred 84% of the segments identified to be similar. The assessment was limited, however, to only five subjects. Extra work can be made to conduct more thorough assessments. Both of these results are nevertheless arguably sufficient to confirm that we have introduced an effective approach to automatic music analysis.

Our recognition work has focused on short segments. It should be straightforward to extend our technique to recognize whole musical scores; one approach is to simply to aggregate the comparisons of individual segments in different pieces. With such extensions, our approach can be applied to areas such as query-by-humming music search engines.

References

[1] J.-J. Aucouturier and F. Pachet. Music similarity measures: What's the use? In *Proc. of the International Conf. on Music Information Retrieval*, 2002.

[2] A. S. Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, MA, 1990.

[3] W. Chai and B. Vercoe. Folk music classification using hidden markov models. In *Proc. International Conf. on Artificial Intelligence*, 2001.

[4] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, 2000.

[5] J. Foote. Content-based retrieval of music and audio. In C.-C. J. Kuo, S. F. Chang, and V. N. Gudivada, editors, *Multimedia Storage and Archiving Systems II, Proc. SPIE*, volume 3229, pages 138–147, 1997.

[6] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith. Query by humming: Musical information retrieval in an audio database. In *Proc. of ACM Multimedia 95*, 1995.

[7] F. Lerdahl and R. Jackendoff. *A generative theory of tonal music*. MIT Press, Cambridge, MA, 1983.

[8] V. I. Levenstein. Binary codes capable of correcting deletions, insertions, and reversals. *Sov. Phys. Dok.*, 10:707–710, 1966.

[9] C. McKay. Automatic genre classification of midi recordings. Master's thesis, McGill University, Canada, 2004.

[10] K. S. Orpen and D. Huron. Measurement of similarity in music: A quantitative approach for non-parametric representations. *Computers in Music Research*, 4:1–44, 1992.

[11] E. Pollastri and G. Simoncelli. Classification of melodies by composer with hidden markov models. In *Proc. of the First International Conf. on WEB Delivering Music*, 2001.

[12] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. In *Proc. on Data Mining and Knowledge Discovery*, volume 2, pages 169–194, 1998.

[13] D. Temperley. *The Cognition of Basic Musical Structures*. MIT Press, Cambridge, MA, 2004.