An Architecture for Effective Music Information Retrieval

Alexandra L. Uitdenbogerd Justin Zobel
Department of Computer Science, RMIT University
GPO Box 2476V, Melbourne 3001, Australia
{alu,jz}@cs.rmit.edu.au

Abstract

We have explored methods for music information retrieval for polyphonic music stored in the MIDI format. These methods use a query, expressed as a series of notes that are intended to represent a melody or theme, to identify similar pieces. Our work has shown that a three-phase architecture is appropriate for this task, in which the first phase is melody extraction, the second is standardisation, and the third is query-to-melody matching. We have investigated and systematically compared algorithms for each of these phases. To ensure that our results are robust, we have applied methodologies that are derived from text information retrieval: we developed test collections and compared different ways of acquiring test queries and relevance judgements. In this paper we review this program of work, compare to other approaches to music information retrieval, and identify outstanding issues.

Keywords: music retrieval system, relevance assessment, music similarity, dynamic programming, string matching, n-grams

1 Introduction

Many people find that they can't identify a piece of music of which they can only remember a fragment. A solution is to visit a music shop and ask the person behind the counter. Alternatively, a music library allows the use of a directory of musical themes to solve the problem — provided the music was written before 1975 [25]. An approach that is becoming more common is to pose the question in a newsgroup on the internet, including any contextual information such as lyrics and where the piece of music was heard, and may even include an audio file containing a recorded or sung portion of the piece. These people are engaging in content-based music information retrieval (MIR).

People trying to find the name of a piece of music are not the only potential users of MIR technology. For example, composers and songwriters may question where their inspiration has come from, forensic musicologists analyse songs for copyright infringement lawsuits, and musicians are often interested in finding alternative arrangements or performances of a particular piece.

Simple MIR technology is in use in practice. Some prototype MIR systems allow melody queries to be posed and present a collection of likely pieces as answers. Most of these search a monophonic database, that is, music in which only one note is sounded at a time. It is not clear how effective these systems are at finding suitable answers, or whether it is possible to achieve good answers to queries presented to a polyphonic database containing a wide variety of musical styles.

We are the principal investigators in an ongoing MIRT project at RMIT to explore technologies for MIR. Our aim in the project is to build a complete architecture for MIR based on the MIDI format, and ultimately to extend the architecture to recorded music. We envisage a system in which a played melody or fragment of recorded music is used as a query and pieces that are thematically similar are returned as matches.

This architecture has several major parts: transformation of arbitrary MIDI files into a form that is suitable for similarity matching and retrieval; development and evaluation of matching algorithms; interfaces for appropriate presentation of matches; and development of test corpuses and methodologies to allow us to reliably distinguish between algorithms.

There have been many matching techniques proposed for MIR prior to our work, but they had not been tested in a common framework, nor had there been any rigorous measurement of effectiveness. Given our background in information retrieval, a primary concern we brought to this work was that our results not have these limitations. At each stage we have aimed to test a wide selection of existing and novel techniques, and have evaluated them, as far as our resources permitted, with independent relevance assessments. Thus one side-effect of the project has been the production of a substantial corpus of multi-channel MIDI files with sets of queries and relevance judgements.

Specific techniques we have tested and developed include:

- *Melody extraction* yielding single-note sequences from arbitrary MIDI data,
- *Melody standardisation* to allow matching across keys and scales,
- *Similarity measurement* primarily based on string matching, and

- *Relevance assessment* based on an audio Web interface.

The project began in 1997, building on Uitdenbogerd's experience as a professional musician and computer scientist, and on Zobel's experience in information retrieval and databases. Other academics that have contributed include Abhijit Chattaraj and Hugh Williams. There are several publications describing our work [28, 29, 30, 31].

2 Issues for MIR

We approached the problem of MIR by adapting ideas from classic information retrieval. However, content-based MIR has distinct differences to other information retrieval tasks, leading to the need for a different approach to processing music queries. There are several factors that make the task difficult:

- Music is polyphonic, that is, more than one note may be sounded simultaneously.
- A query may be presented in a different key to the stored version.
- There are likely to be variations between the query and answers such as repeated notes, ornamentation, and unstressed notes with different pitch. (There are similar variations between different performances and arrangements of the same piece.) Pieces of music are usually perceived as highly similar despite such differences in the notes.
- If the query is presented by singing, there are likely to be errors due to limitations in singing accuracy.
- While a query is likely to be based on the melody of a piece of music, it is not clear what the melody of a piece is: it is as obscure a concept as the "meaning" of a piece of text.
- The commonest representations of music are as audio streams, not as sequences of notes.

The choices when developing a prototype MIR system are many. The types of data to work with include audio, polyphonic note data, and monophonic note data. To develop a MIR system that successfully matches a query melody to a polyphonic database of music, several problems must be solved. First, it is necessary to decide what information — that is, what representation of the music — the query should be matched against. Second, an appropriate matching method needs to be developed. Third, it is important to choose an appropriate method of evaluating the effectiveness of techniques developed.

To our knowledge, no researchers, including us, have developed a workable melody matching system based on stored audio. There are developing technologies that may make such matching feasible, such as effective methods for note extraction and key and tempo identification. It is reasonable to suppose that such techniques will in the future allow translation of audio into a note-based representation, thus reducing the matching problem to the issue of comparing a query to a polyphonic stream of notes.

Matching a query melody against polyphonic music is difficult with most sources of polyphonic music data, as the melody component of the music is not defined, and the vast majority of notes in complex music do not contribute to the perceived melody.

This leaves two main approaches: match the query against any possible sequence of notes through each piece of music, or decide which portion of the music the query should be matched against and only match against that. The first approach leads to combinatorial explosion, and is likely to produce many false matches. The second approach involves a "melody extraction" phase, in which the notes that belong to the melody or theme are selected for matching. For some queries, it may be more appropriate to select other portions of the music, such as chordal information.

A central assumption we have made is that queries are likely to be based on the "melody" of a piece of music. We predict that the majority of monophonic queries will be of this type. It is quite possible that a query will be on some other part of the music, such as a bass line. However, techniques that involve matching against the "melody" of each part of a polyphonic piece work well and therefore can satisfy queries on non-melody parts. Our approach can also be used for some forms of polyphonic match, in which case the melody can be automatically extracted from the query before being matched against the collection. A further, more intensive matching process can then be applied to a small pool of best answers.

An approach based on note information does not allow searches on all types of music. First, some music created in this century is not note-based at all, consisting of organised sound instead of melodies. For this type of music, there may not be any melody which could be used as a query. The user may need to use an audio-based technique for successful location of the work — probably by using a sample noise as a query. Second, we do not address music using scales and tunings other than the western scale. This is not a problem for pentatonic music (five note scale) which is easily represented, but microtonal music (music containing intervals that are smaller than a semitone) is not catered for at all. Music intended for different tunings may be represented if the scales contain less than twelve notes per octave, but will not sound as intended and thus may not match with appropriate pieces. However, as contour and relative interval size is important for matching, the disadvantage for differently tuned music may be small.

3 MIR methods developed at RMIT

The MIRT project at RMIT began in 1997. Our primary aims in the project have been to put MIR on as solid a research foundation as classic information retrieval, and to use this foundation to develop novel MIR techniques. Our work, therefore, has two threads: one is collection of test data and development of methods for measuring MIR systems; the other is development of the MIR systems themselves.

We have devised a three-stage approach for MIR consisting of melody extraction, melody standardisation, and similarity measurement. Melody extraction, or TRIMming, involves choosing melody notes out of polyphonic pieces of music to simplify the matching problem and reduce the pool of irrelevant matches. Melody standardisation converts the melodic information into a form consisting of a sequence of symbols. The symbols represent the features used for melodic matching. The similarity measurement stage calculates a similarity score based on the standardised form of the melodies being matched.

Since melody extraction is inaccurate, and since — even with standardisation — the same melody can be represented by different sequences of notes, two strings representing the same melody are not necessarily identical. Matching must therefore be based on some measure of the similarity of query and piece. Given a standardised query string and a collection of standardised piece strings, matching involves computing a numerical score for each piece with respect to the query. Like other IR systems, the pieces can then be sorted by their score, and the highest-ranked pieces returned to the user as potential matches.

Music perception research provides us with some clues as to what would be perceived as melody. Typically, the highest-pitch notes would be classed as the melody notes unless they are monotonous. For other notes to be identified as the melody, various compositional tricks need to be applied such as making it much louder than the accompanying notes and making use of a single timbre throughout the melodic phrase. Sometimes the pitch of soloists is slightly higher than normal to accentuate the melody against a large body of sound [14].

Findings from music perception research were a useful guide in deciding which techniques might be effective at MIR. In particular, the research into how melody is perceived has informed our approach to melody extraction. Listeners usually hear the highest pitch notes as melody notes, except when they are monotonous [12]. Pitch proximity is the most important factor in grouping notes into perceived parts [4]. These concepts have been incorporated into our melody extraction techniques and tested with human listeners. The results of our experiment on melody extraction confirm the results from music psychology, in that extracted notes that consisted of the highest pitch at any instance were deemed to be most like the melody of the pieces listened to.

The importance of melodic contour as a feature of both musical memory [6] and singers' accuracy [19] makes it a feature that ideally should be included in any melody standardisation process. We have chosen standardisation methods that retain contour information for our matching experiments.

In the final stage of our approach we applied matching techniques, namely, dynamic programming and n-gram scoring, to melody matching. Different variations on the above two methods were tested to determine retrieval effectiveness.

Our experiments revealed that our three-stage approach can successfully answer melody queries. A technique that selected the highest pitch notes commencing at each instant, which we call "all-mono", was the most effective melody extraction technique as judged by listeners. It was also shown to be effective in melody matching. Melody standardisation techniques tested included contour and interval-based techniques. These use a relative pitch approach, allowing matching in any key. Contour standardisation reduces the melody to a string of characters representing "up", "down", and "same" pitch directions. This was shown to be insufficient with our collection for queries of even 20 notes. Our experiments showed that interval-based techniques are vastly more effective. Matching techniques based on either local alignment dynamic programming or n-grams were shown to be the best at retrieving useful answers.

4 Methodology and measurement

In order to evaluate the effectiveness of different melody matching techniques, we collected query and relevance sets in two ways. Query sets consisted of melodies that were automatically extracted from the pieces, and truncated to specific lengths. The second type of query set was created by obtaining manual queries via a volunteer who listened to pieces and played a representative melody fragment. To form relevance judgements, we first found pieces in the collection for which there was more than one version, using the filenames and listening to verify that it was indeed the same piece. This formed one relevance set. Relevance judgements were also collected from users who listened and judged how similar pieces were to each manual query.

These query and relevance sets were used to measure the quality of the answers retrieved by the different matching methods. As the same pieces were used for automatic and manual queries, we had both automatic and manual judgements in each case, allowing full evaluation of each method of measuring system performance. Using these sets of queries and relevance judgements, we could measure performance. We used two standard IR measures of retrieval effectiveness: eleven-point precision averages, and precision at ten.

The process of collecting relevance judgements revealed some issues that need to be resolved for MIR user interfaces. Some tasks are very difficult and time-consuming for users, such as comparing two unfamiliar pieces of music.

In addition to evaluating the techniques for melody matching, we tested the evaluation methodology itself, by comparing the effect of the two sets of queries and relevance judgements. We found that the two query sets gave significantly different results when used to rank melody matching techniques. The two sets of relevance judgements were more consistent but had some minor differences.

5 Other MIR projects

MIR technology was much less mature when this project began. In some respects, our methods provide a consistent framework in which other MIR work can be placed, and it is therefore interesting to review this work in the context of our proposals.

Some MIR research has concentrated on monophonic information and others on polyphonic. Most research that uses monophonic musical information has as its source of data a collection of folk songs [8, 22, 24, 26], or a small collection of melodies [15]. A variety of techniques have been used to locate melodic matches, including dynamic programming [22, 23], n-gram-based matching [9], feature histograms [17], and state matching [22].

Some researchers have chosen to match a query melody against melodies automatically extracted from polyphonic musical data [1, 13]. Simple heuristics were used to determine the melody but there was no evaluation of the techniques. Very little has been published on the process of melody extraction. There have been attempts at splitting polyphonic music into parts with moderate success [20], but the melody extraction problem has not previously been explored in the context of retrieval. Researchers who have chosen polyphonic note data for their research [1, 5, 13, 18] use collections of files of data in the MIDI file format as their source. These collections not only provide polyphonic musical data, but include music in a wide variety of styles.

Other researchers have chosen to match queries against any occurrences within a collection, regardless of whether these are across musical parts or instruments [3, 5, 18], however, this approach has not yet been shown to produce effective answers, and based on our own (unpublished) experiments, would require the use of more detailed queries to produce good results.

Some researchers have tackled the problem of audio retrieval, in which similarity is determined by extracting features from wave-forms. The state of the art in this area is that it is possible to determine whether a wave-form contains music or speech, some stylistic aspects can be detected in musical recordings, the beats and therefore the tempo can be determined, but identification of actual notes within a non-monophonic piece of music is very difficult. The main problem is that it is hard to distinguish between harmonics and notes. For example if the note A at 220 Hz is played on a piano, the wave-form of that note includes harmonics at double (440 Hz), triple (660 Hz), quadruple (880 Hz) frequency and so on. Note identification is a little easier if constraints are placed on the music to be analysed, such as using an instrument with a simple timbre so that features of the instrument's typical sound pattern can be used to make sense of the wave-form. However, even so the process is not accurate. The focus of research in this field is the identification of new techniques for musical transcription from audio waves [27] and use of non-melodic features for matching, such as structural information [10, 11].

Few MIR techniques have been evaluated for effectiveness using standard information retrieval methodology. Some matching techniques are applied to a small set of pieces and are subjectively evaluated by the researchers [23, 24]. Others have used statistical measures of success [1, 7, 22], or known item searches [9, 15]. The best evaluation technique that has been applied was the testing of a set of 100 hummed queries against a set of 500 songs. The number of pieces that were retrieved in the top 10 and top 1 were reported for four different algorithms [15, 16].

6 Conclusions

Using our test corpus, we have concluded that simple melody extraction and melody standardisation is sufficiently effective to be used in practice. Among the similarity techniques tested, both approximate string matching and n-gram matching have strengths and weaknesses, but both are reliable at identifying pieces of music that are similar to a query presented as single-note MIDI data. We have also found that the kind of matching technique that works best depends on the task: matching two pieces of music is a different problem to that of matching a query to a piece of music.

It would not be easy with current music processing technology to extend the standardisation and matching techniques to data in an audio format. Also, interface design for relevance assessment is not straightforward, because of the time required to manually compare two pieces of music. It is clear from our work in this area — and in particular from our work's shortcomings — that good visualisation is essential to rapid relevance assessment.

Overall, we believe that it would not have been possible to draw strong conclusions (or, possibly, any conclusions at all) without a test corpus and relevance judgements. Such a resource is essential to work in this area.

Acknowledgements

This work was supported by the Australian Research Council.

References

[1] S. Blackburn and D. De Roure. A tool for content-based navigation of music. In *Proc. ACM International Multimedia Conference*. ACM, September 1998.

[2] D. Byrd, J. S. Downie, T. Crawford, W. B. Croft, and C. Nevill-Manning, editors. *International Symposium on Music Information Retrieval*, volume 1, Plymouth, Massachusetts, October 2000.

[3] M. Clausen, R. Engelbrecht, D. Meyer, and J. Schmitz. PROMS: A web-based tool for searching in polyphonic music. In Byrd et al. [2].

[4] D. Deutsch. Grouping mechanisms in music. In D. Deutsch, editor, *The Psychology of Music*, chapter 4, pages 99–134. Academic Press, Inc., 1982.

[5] M.J. Dovey. An algorithm for locating polyphonic phrases within a polyphonic musical piece. In *Proceedings of the AISB Symposium on Musical Creativity*, Edinburgh, April 1999. AISB.

[6] W.J. Dowling. Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85(4):341–354, 1978.

[7] J.S. Downie. Informetrics and music information retrieval: an informetric examination of a folksong database. In *Proceedings of the Canadian Association for Information Science, 1998 Annual Conference*, Ottawa, Ontario, 1998. CAIS.

[8] J.S. Downie. *Evaluating a Simple Approach to Musical Information Retrieval: Conceiving Melodic N-grams as Text.* PhD thesis, University of Western Ontario, 1999.

[9] J.S. Downie. Access to music information: The state of the art. *Bulletin of the American Society for Information Science*, 26(5), June/July 2000.

[10] J. Foote. Visualizing music and audio using self-similarity. In *Proc. ACM International Multimedia Conference*, pages 77–80, Orlando Florida, USA, October 1999.

[11] J. Foote. ARTHUR: Retrieving orchestral music by long-term structure. In Byrd et al. [2].

[12] R. Francès. *La Perception de la Musique.* L. Erlbaum, Hillsdale, New Jersey, 1958. Translated by W.J. Dowling (1988).

[13] A. Ghias, J. Logan, D. Chamberlin, and B. Smith. Query by humming — musical information retrieval in an audio database. In *Proc. ACM International Multimedia Conference*, 1995.

[14] S. Handel. *Listening: An introduction to the perception of auditory events.* MIT Press, 1989.

[15] T. Kageyama, K. Mochizuki, and Y. Takashima. Melody retrieval with humming. In *Proc. International Computer Music Conference*, 1993.

[16] T. Kageyama and Y. Takashima. A melody retrieval method with hummed melody. *Transactions of the Institute of Electronics, Information and Communication Engineers*, J77-D-II(8):1543–1551, 8 1994. in Japanese.

[17] N. Kosugi, Y. Nishihara, S. Kon'ya, M. Yamamuro, and K. Kushima. Music retrieval by humming - using similarity retrieval over high dimensional feature vector space. In *Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, Victoria, Canada, August 1999.

[18] K. Lemström and J. Tarhio. Detecting monophonic patterns within polyphonic sources. In J. Mariani and D. Harman, editors, *Proc. Conference on Content-Based Multimedia Information Access*, volume 6 of *RIAO*, Paris, France, April 2000.

[19] A. T. Lindsay. Using contour as a mid-level representation of melody. Master's thesis, MIT, Massachusetts, 1996.

[20] A. Marsden. Modelling the perception of musical voices: a case study in rule-based systems. In Marsden and Pople [21], pages 239–263.

[21] A. Marsden and A. Pople, editors. *Computer Representations and Models in Music.* Academic Press, London, England, 1992.

[22] R.J. McNab, L.A. Smith, I.H. Witten, C.L. Henderson, and S.J. Cunningham. Towards the digital music library: Tune retrieval from acoustic input. In *Proc. ACM Digital Libraries*, 1996.

[23] M. Mongeau and D. Sankoff. Comparison of musical sequences. *Computers and the Humanities*, 24:161–175, 1990.

[24] D. O'Maidin. A geometrical algorithm for melodic difference. *Computing in Musicology*, 11:65–72, 1998.

[25] Parsons. *The Directory of Tunes.* Spencer Brown and Co., Cambridge, England, 1975.

[26] H. Schaffrath. The retrieval of monophonic melodies and their variants: Concepts and strategies for computer-aided analysis. In Marsden and Pople [21], pages 95–110.

[27] E.D. Scheirer. *Music Listening Systems.* PhD thesis, MIT, Massachusetts, June 2000.

[28] A. L. Uitdenbogerd. *Music Information Retrieval Technology.* PhD thesis, School of Computer Science and Information Technology, RMIT University. In submission.

[29] A. L. Uitdenbogerd and J. Zobel. Manipulation of music for melody matching. In B. Smith and W. Effelsberg, editors, *Proc. ACM International Multimedia Conference*, pages 235–240, Bristol, UK, September 1998.

[30] A. L. Uitdenbogerd and J. Zobel. Melodic matching techniques for large music databases. In D. Bulterman, K. Jeffay, and H. J. Zhang, editors, *Proc. ACM International Multimedia Conference*, pages 57–66, Orlando, Florida, November 1999.

[31] A. L. Uitdenbogerd and J. Zobel. Music ranking techniques evaluated. In *International Symposium on Music Information Retrieval*, Plymouth, Massachussetts, USA, October 2000. poster.