

Melody Retrieval On The Web

Thesis proposal for the degree of Master of Science
at the Massachusetts Institute of Technology

Wei Chai

M.I.T Media Laboratory

Fall 2000

Thesis supervisor:

Barry Vercoe
Professor, M.I.T Media Laboratory

Thesis reader:

Joseph A. Paradiso
Principal Research Scientist, M.I.T Media Laboratory

Thesis reader:

Christopher Schmandt
Principal Research Scientist, M.I.T Media Laboratory

Abstract

The emergence of digital music on the Internet requires new information retrieval methods adapted to specific characteristics and needs. While music retrieval based on the text information, such as title, composers, or subject classification, has been implemented in many existing systems, retrieval of a piece of music based on musical content, especially an incomplete, imperfect recall of a fragment of the music, has not yet been fully explored.

This thesis is to explore the main problems involved in a web-based melody retrieval system. I propose to build a query-by-humming system, which can find a piece of music in the digital music repository based on a few hummed notes, using a melody representation that combines both the pitch contour and the beat information. Since an input query (hummed melody) may have various errors due to uncertainty of the user's memory or the user's singing ability, the system should be able to tolerate the errors. Furthermore, extracting melodies to build a melody database is also a complicated task. Therefore, melody representation, query construction, melody matching and melody extraction are critical for an efficient and robust query-by-humming system. Thus, they are the main problems to be solved in the thesis.

Contents

1	Introduction	1
2	Background.....	1
2.1	Related Work	1
2.2	MPEG7 Standard	1
3	Scope	2
3.1	Assumptions.....	2
3.2	Objectives	2
3.3	System Design	3
3.4	Major Problems.....	4
4	Approach	4
4.1	Melody Extraction	4
4.2	Melody Representation	4
4.3	Query Construction.....	5
4.4	Melody Matching.....	5
5	Deliverables	6
6	Logistics	6
6.1	Schedule.....	6
6.2	Resources	6
	References	7

1 Introduction

The Internet has become an important source for people to obtain multimedia content for entertainment, education, business, or other purposes. The emergence of digital music on the Internet requires new information retrieval methods adapted to specific characteristics and needs. Although there are lots of websites for music sale, advertisement and sharing, the interfaces are not so convenient for finding desired music: only category-based browsing and/or text-based searching are supported. To find a piece of music, the user needs to know the title, composer, artist or other text information, so that the search engine can search the database based on that information; otherwise the user needs to browse the whole category. This procedure might be very time-consuming. Therefore, developing new methods to help people retrieve music on the Internet is of great value.

My thesis is to build a query-by-humming system (called the QBH system), which can find a piece of music in the digital music repository based on a few hummed notes. So when the user does not know the title or any other text information about the music, he is still able to search for music by humming the melody. It's a much friendlier interface for the Internet music searching.

2 Background

2.1 Related Work

The query-by-humming system was first proposed by Ghias *et al.* (1995). Following Ghias *et al.*, several research groups including the New Zealand Digital Library project by McNab *et al.* (1996), the Search By Humming project by Blackburn *et al.* (1998), the Themefinder project by Stanford University, the TuneServer project by University of Karlsruhe, the MiDiLiB project by University of Bonn, etc. are working in this area. Up to now, most projects use music corpora in symbolic representation, such as MIDI, to build the melody database. Three-level pitch contour melody representation (U/D/S indicating that the interval goes up, down or remains the same) is widely adopted. Some systems support waveform queries (e.g. humming or whistling), while some other systems only support text format queries. Various approximate matching methods are used.

2.2 MPEG7 Standard

MPEG7, also called Multimedia Content Description Interface, is a standard for describing the multimedia content data to allow universal indexing, retrieval, filtering,

control, and other activities supported by rich meta-data. For music, a set of descriptors and description schemes that can best describe music for retrieval purpose will be adopted. The melody representation that we use in the QBH system has been proposed to the MPEG7 standard committees. By building this system, we can testify whether the representation is robust, efficient and concise for inclusion in MPEG7.

3 Scope

3.1 Assumptions

Defining what exactly is or is not a melody can be somewhat arbitrary. Melodies can be monophonic, homophonic, or contrapuntal. Sometimes what one person perceives to be the melody is not what another perceives. A melody can be pitched or purely rhythmic, such as a percussion riff. Our research does not attempt to address all of these cases and is limited in scope to pitched, monophonic melodies. We assume that all the pieces in our music corpus have monophonic melodies, which the users can easily and consistently identify. Furthermore, the user should be fairly familiar with the melody he wants to query, though perfect singing skill is not required.

3.2 Objectives

Besides its application value, the query-by-humming system is also an interesting topic from a scientific point of view. Identifying a musical work from a melodic fragment is a task that most people are able to accomplish with relative ease. However, how people achieve this is still unclear, i.e. how do people extract melody from a complex music piece and convert it to a representation that could be memorized and retrieved easily? Although this whole question is beyond the scope of this thesis, we will build a system that performs like a human: it can “read” scores or even “hear” music to extract melodies; convert the melodies into an efficient representation and store them in its “memory”; when a user asks for a piece of music by humming the melody, it can first “hear” the query and then search in its “memory” for the piece that it “thinks” most similar to the query. Thus, the research I propose will explore the melody retrieval problem from two perspectives: as a practical solution to a query-by-humming system, and as a scientific inquiry into the nature of the melody perception process.

The main, aware features of this system as compared with other existing systems are:

- (1) A new melody representation, which combines both pitch and beat information.
- (2) A new matching algorithm based on the representation.
- (3) A possible method for extracting melodies from polyphonic tracks for building a melody database.

3.3 System Design

The QBH system adopts the client-server architecture and consists of four parts (Figure 1):

- (1) Music database: It includes the source data and the target data. The source data are the original music corpus, from which we extract melodies and generate the target data. They are in various symbolic representations (i.e. scores and MIDIs). The target data are the data that the end user can play back at the client side. The target data in the system are in MIDI format.
- (2) Melody description object: It is a binary persistent object that capsulates the melody information based on our melody representation and acts as an efficient indexing of the music corpus.
- (3) QBH Server: It receives the query from the QBH client; matches it with the melodies in the melody description object; and retrieves the target data with the highest matching scores.
- (4) QBH client: It tracks the pitch contour and timing information from the user's humming signal; constructs the query; sends the query to the QBH server via CGI; receives the result and plays it back.

Besides the above four parts, we need to develop some tools to build the whole system. They include

- A score to midi tool
It converts score files (in .krn, .alm, .esc and .niff formats) into midi files.
- A melody extraction tool
It extracts the melody information and constructs the melody description object.

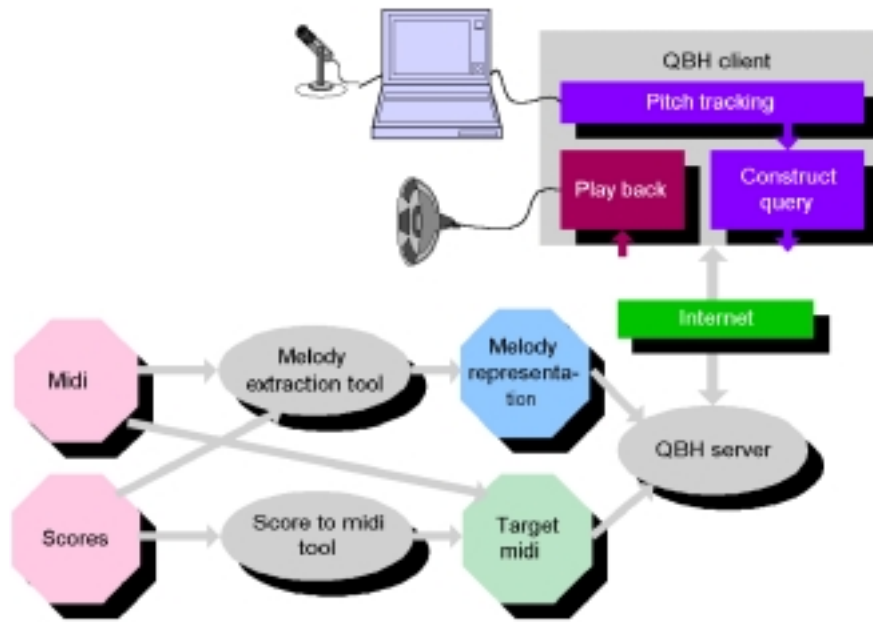


Figure 1: QBH system architecture

3.4 Major Problems

The main issue in building such a melody retrieval system is that an input query (hummed melody) may have various errors due to uncertainty of the user's memory or the user's singing ability. To tolerate the errors, we need effective representation and reasonable approximate matching method. Besides, extracting melody information from existing music corpus is not a trivial task. Therefore, I divide the whole problem into four sub-problems: the melody extraction problem, the melody representation problem, the query construction problem and the melody matching problem. I believe the above four problems are the key points of a complete, robust and efficient query-by-humming system, and thus the focuses of my thesis.

4 Approach

4.1 Melody Extraction

What is melody and how do humans perceive melody from a complex piece of music? Is it possible to extract melodies from existing digital music corpora by machine and use them to build a melody database for the retrieval purpose? There could be two kinds of sources from which we extract melodies and build our melody database: corpora in symbolic representation such as score formats and MIDI; corpora in waveform representation. In this system, I'll only use symbolic corpora, because extracting melody from waveform representation involves the automatic transcription problem, which has not yet been solved well. Extracting melody information from the symbolic representation with a separate monophonic melody track is relatively easy. I've already developed a tool to manipulate this kind of corpora. However, in many cases, a melody is contained in a polyphonic track. Extracting melody information from such a polyphonic track is quite hard. Uitdenbogerd (1998) proposed an algorithm, sometimes called the "skyline algorithm", which basically extracts the highest pitch line as the melody line. It turns out to be effective for some music genres like pop music, but doesn't work well for classical music, etc. I've modified the skyline algorithm by adding one parameter called time overlap parameter, which can make the recognized melody more accurate than the original algorithm on the test set provided by Uitdenbogerd. I'm also exploring another algorithm using clustering method to group notes into separate lines according to their pitch and time relations.

4.2 Melody Representation

What are the significant features people use to identify a melody or to distinguish between melodies? How can the melodies be represented sufficiently as well as concisely? Previous work mostly proposes using pitch contours to represent melodies.

They seldom use rhythm in their melody representation. However, rhythm is obviously important, because when identifying a melody, the listener perceives not only the pitch/interval information in the melody, but also how those notes correspond to particular moments in time. Rhythm is one dimension in which melodies in general can not be transformed intactly (Kim *et al.*, 2000). A representation combining both pitch and rhythm information, which we call TPB representation, was proposed by Kim *et al.* (2000) and adopted in this prototype system. Its performance will be compared with the other widely used representations.

4.3 Query Construction

Query construction means how we can obtain the pitch contour and rhythm information from the hummed query. Real-time pitch tracking algorithms will be implemented in this system. Human voice is the stimuli to be tracked, and sound segmentation is very important for the robustness of the system. I've implemented the algorithm that is also widely used in query-by-humming systems, using amplitude to segment the notes and autocorrelation to do the pitch tracking. Several other algorithms will also be explored and compared (Rabiner, L.R. *et al.*, 1976). To obtain rhythm from user's query is more complicated. I thus implemented an interface that can click the rhythm based on the time signature and tempo the user input. When the user hums the melody with the clicking signal, we can separately record the beat information of the query. This interface turns out to be quite effective when the user is familiar with the melody and has the basic sense of rhythm. I'm also exploring an algorithm to obtain the rhythm from the query hummed without clicking, which will give the user more freedom. Scheirer (2000) proposed a beat tracking algorithm for raw audio music. We can use a similar method to achieve our goal.

4.4 Melody Matching

How do people measure the similarity of melody? Or, how can we retrieve melody from our mind so easily even after it is transformed in some way? This problem is closely related to the representation problem. Levitin (1999) described melody as an "auditory object that maintains its identity under certain transformations ... along the six dimensions of pitch, tempo, timbre, loudness, spatial location, and reverberant environment; sometimes with changes in rhythm; but rarely with changes in contour". Although rather broad, this definition highlights several important properties of melody. People are able to recognize melodies even when they are played on different instruments, at different volumes, and at different tempi (within a reasonable range). Based on the TPB representation, we also proposed an approximate string matching method to do the task (Kim *et al.*, 2000). This algorithm takes into account not only robustness but also efficiency.

5 Deliverables

I intend to provide the following results:

- An efficient and concise melody representation scheme.
- A robust melody matching algorithm based on the above melody representation.
- An implementation of the QBH system that includes the client, the server and a melody database.
- A set of ancillary tools for building the QBH system, e.g. the melody extraction tool, the score-to-midi tool, etc.
- Experimental results comparing our system with other existing systems in terms of the representation, the algorithms and the system implementation.

6 Logistics

6.1 Schedule

Up to now, I've built a melody database and implemented a client-server structured prototype system. For my thesis work, I still need to

- develop new melody extraction algorithms and enlarge the current melody database;
- test our matching algorithm based on our proposed representation and compare it with other algorithms;
- improve the robustness of the pitch tracking and make the interface friendlier;
- evaluate the system. I'll put the system online when it is ready and analyze the user's query records in terms of efficiency and accuracy.

I plan to finish these tasks within four months.

6.2 Resources

Here is a list of resources required to carry out this work:

- A PC machine running windows NT.
- A PC machine (or laptop) running windows 98/2000 with a duplex sound card.
- Software including MS VC++, Matlab with mcc compiler and math library.
- Music data in symbolic representation.

References

- Blackburn, S. and DeRoure, D. "A tool for content based navigation of music." Proc. ACM Multimedia, Bristol, 1998. <http://audio.ecs.soton.ac.uk/sbh>.
- Chai, Wei and Vercoe, Barry. "Using user models in music information retrieval systems." Proc. International Symposium on Music Information Retrieval, Oct. 2000.
- Chen, J. C. C. and Chen, A. L. P. "Query by rhythm: an approach for song retrieval in music databases" Proc. Eighth International Workshop on Research Issues In Data Engineering, 1998.
- Chou, T. C.; Chen, A. L. P. and Liu, C. C. "Music database: indexing techniques and implementation." Proc. International Workshop on Multimedia Database Management Systems, 1996.
- Dowling, W. J. "Scale and contour: Two components of a theory of memory for melodies." Psychological Review, vol. 85, no. 4, pp. 341-354, 1978.
- Ghias, A.; Logan, J.; Chamberlin, D. and Smith, B. C. "query by Humming: musical information retrieval in an audio database." Proc. ACM Multimedia, San Francisco, 1995.
- Kim, Youngmoo; Chai, Wei; Garcia, Ricardo and Vercoe, Barry. "Analysis of a contour-based representation for melody". Proc. International Symposium on Music Information Retrieval, Oct. 2000.
- Kosugi, N.; Nishihara, Y.; Kon'ya, S.; Yamamuro, M. and Kushima, K. "Music retrieval by humming-using similarity retrieval over high dimensional feature vector space." Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 1999.
- Levitin, D. J. "Memory for Musical Attributes" from Music, Cognition, and computerized Sound, ed. Perry R. Cook. Cambridge, MA: MIT Press, 1999, 214-215.
- Lindsay, A. T. Using contour as a mid-level representation of melody. MS thesis. MIT Media Lab, 1996.
- Liu, C. C.; Hsu, J. L. and Chen, A. L. P. "Efficient theme and non-trivial repeating pattern discovering in music databases", Proc. 15th International Conference on Data Engineering, 1999.
- McNab, R. J.; Smith, L. A.; Witten, I. H.; Henderson, C. L. and Sunningham, S. J. "Toward the digital music library: tune retrieval from acoustic input." Proc. ACM Digital Libraries, Bethesda, 1996. <http://www.nzdl.org>.

MiDiLiB, University of Bonn, <http://leon.cs.uni-bonn.de/forschungprojekte/midilib/english>.

Pollastri, E. “Melody-retrieval based on pitch-tracking and string-matching methods.” Proc. Colloquium on Musical Informatics, Gorizia, 1998 .

Rabiner, L. R.; Cheng, M. J.; Rosenberg, A. E. and McGonegal, C. A. “A comparative performance study of several pitch detection algorithms”, IEEE Trans. on Acoustics, Speech and Signal Processing, vol ASSP-24, no.5, 1976, 399-418.

Scheirer, E. D. Music-Listening Systems. PhD thesis. MIT Media Lab, 2000.

Themefinder, Stanford University, <http://www.ccarh.org/themefinder>.

Tseng, Y. H. “Content-based retrieval for music collections.” Proc. Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1999.

TuneServer, University of Karlsruhe, <http://wwwipd.ira.uka.de/tuneserver>.

Uitdenbogerd, A. L. and Zobel, J. “Manipulation of music for melody matching.” Proc. ACM International Conference on Multimedia, 1998.