A Hierarchical Approach: Query Large Music Database by Acoustic Input

Yazhong Feng   Yueting Zhuang   Yunhe Pan
Department of Computer Science, Zhejiang University, Hangzhou 310027, China

Categories & Subject Descriptors: H.3.3 Retrieval models, H.3.7 Systems issues

General Terms: Algorithms, Design

Keywords: music information retrieval, recurrent neural network,correlation degree

1. INTRODUCTION

Digital music download activity is becoming the dominant traffic stream on internet, research on content-based music retrieval tool is increasing. Our goal is to develop efficient music content search engine which accepts multi-modal inputs, such as score, text, speech, and humming, query by score and query by text such as music metadata or lyrics are traditional music retrieval methods, query by speech is a query by text application in nature, except that the speech to text interface is employed, query by humming (QBH) is the most natural way to query music by its content. A rich range of researchers are contributing to content-based music retrieval, [1] introduced a "Query by Humming" system, [2] developed a system to query music in digital library via acoustic input, [3] retrievals music by text retrieval, [4] retrievals music by beat information, [5] made index on music database by recurrent neural network. With regard to melody representation, only pitch information is used in [1], [6] represents melody by a triple <T P B>, [7] uses another triple (pitch contour, pitch interval, duration) to represent melody. We implement a music retrieval system on a large music database, which contains MIDI music, it accepts acoustic input, indices are made on music database via the weights of recurrent neural network, and retrieval results are ranked by self-designed correlation degree.

2. MUSIC REPRESENTATION

To retrieval music by acoustic input, we represent both human humming and music in database by <P B>, P is pitch contour, B is beat information. We know that the most impressed part of music is its melody, pitch and rhythm are important components of melody, pitch determines which note is played by instruments or sung by people, to eliminate transcription or memory errors, we use pitch contour for pitch information, rhythm is the beat movement, we represent rhythm by beat in this paper. The reason why we use only pitch contour and beat information to represent melody is as followings: when humming, people tend to use his or her own tempo, which is sometime different from original music, so we do not use tempo information such as T in <T P B> of [6], most people can not keep right pitch interval and duration when they hum, so, unlike [7], we use only pitch contour for pitch information. Autocorrelation or cepstrum is used to track pitch from acoustic input, beat information is derived from acoustic input by method suggested in [9].

SIGIR'02, August 11-15, 2002, Tampere, Finland.
ACM 1-58113-561-0/02/0008.

3. HEURISTIC RULES

To make music retrieval more robust, music perception knowledge should be embedded into signal processing tools to perform pitch tracking of humming, after analyzing our music database which has about 1200 pieces of MIDI music, we deduce the following heuristic rules to aid signal processing in music retrieval:

- Five-level pitch contour is more robust than three-level contour in music matching.

Based on our music database, the statistic result is almost the same as [4][6], the melodic intervals of most adjacent MIDI notes are in the range of [-2 2].

The five-level contour (-2 –1 0 1 2) is defined as following:

$$Contour_5 = \begin{cases} -2, when N_{i+1} - N_i < -2 \\ -1, when -2 \le N_{i+1} - N_i < 0 \\ 0, when N_{i+1} = N_i \\ 1, when 0 < N_{i+1} - N_i \le 2 \\ 2, when 2 < N_{i+1} - N_i \end{cases}$$

N_i is the i th MIDI note.

- Frequency of human humming or singing is in the range of [80Hz, 800Hz]. When performing pitch tracking, if the fundamental frequency of a frame is not in above range, this frame is not a note candidate.
- The melody variation of humming or singing is not very severe. If the fundamental frequency of a frame is bigger than that of each two frames on its left and right, this frame is not a note candidate.
- The experimental knowledge [4] that most music is self-similar supports the use of recurrent neural network to remember melody and make index.

4. SYSTEM COMPONENTS

Our system has three components: signal processing of humming input, music database and matching strategy.

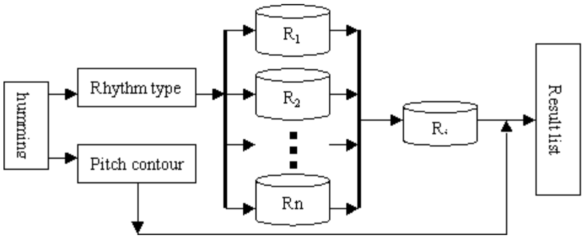

Figure 1. Components of our MIR system.

Pitch-tracker translates humming into Five-level pitch contour, rhythm detector calculate time signature from humming. Music database is separated into regions according to rhythm type, such as

T=(2/4, 3/4, 5/4, 6/4, 7/4, 6/8). Let M be a music database, $n = |T|$, then $\mathrm{M} = \bigcup_{i=1}^{n} R_i$.

4.1 Pitch Tracker

Although pitch tracking on polyphonic music is not very successful, tracking pitch from acoustic humming input is straightforward and reliable; usually autocorrelation or cepstrum is employed to track pitch.

4.2 Rhythm Detector

Our rhythm detector is the same as [8], MFCC (Mel-Frequency Cepstral Coefficients) feature vector v is extracted from humming audio, and the element of similarity matrix S is calculated by

$$s(i,j) = \frac{v_i \bullet v_j}{\|v_i\|\|v_j\|} . \qquad \textbf{(Eq. 1)}$$

Beat spectrum (Eq.2) reveals that the beat information of humming input is b :

$$B(k,l) = \sum_{i,j} S(i,j)S(i+k, j+l) \qquad \textbf{(Eq. 2)}$$

4.3 Music Database Regions

Beat information b is used to locate the region in which to match humming with music, let $i = \arg(\min(|b-T|))$, then R_i is the selected region.

4.4 Indices on Music Database

Melody contour for each music is trained by a recurrent neural network [5], its weight matrix acts as the index of this piece of music in the database. Because of the same structure of neural network, index size for every piece of music is same and the indices on the whole music database increase linearly to the database size.

4.5 Retrieval by Correlation Degree

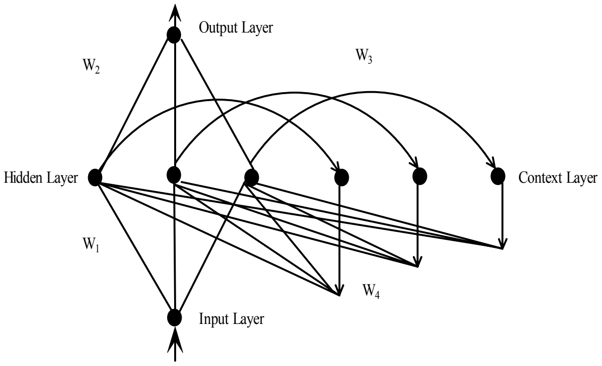

Figure 2. Structure of recurrent neural network used for memorizing music.

Let P be pitch contour of humming, then the retrieval result is $\arg(\max(P\aleph R_i))$ .Let $R_i = \{m_{ij}\}$, where m_{ij} be the j th piece of music in region R_i , $Net(m_{ij})$ is the recurrent neural network that remembers music m_{ij} , $P * Net(m_{ij})$ is the network output when input is P , define $\gamma(P, P*Net(m_{ij}))$ to be the correlation degree of P with $P*Net(m_{ij})$, then

$$P\aleph R_i = \{\gamma(P, P*Net(m_{ij}))\}, j = |R_i| \qquad \textbf{(Eq. 3)}$$

5. Experimental Result

Experiment on the database that has 1200 pieces of MIDI music shows the following result:

Resolution	Successful Rate
Top3	68%
Top10	89%

It is much more difficult to extract melody from polyphonic music in raw audio format such as WAV or AU, otherwise different approach that matches singing which is characteristic of songs may be promising. We are now exploring a technique to retrieval popular songs directly by singing with lyrics, the basic idea is to extract singing from stereo music recordings by independent component analysis (ICA) [10], then we perform matching on melody contour derived from singing. We hope this will contribute a bit to multi-modal model of music information retrieval system.

6. REFERENCES

[1] Ghias, A. Logan, J. Chamberlin, D. and Smith, B. "Query by humming," in ACM Multimedia, 1995.

[2] McNab, R. Smith, L. Witten, I. Henderson, C. and Cunningham, S. "Towards the digital music library: Tune retrieval from acoustic input," in Digital Libraries 1996.

[3] Downie, J. S. "Music retrieval as text retrieval: Simple yet effective," in Proceedings of SIGIR '99 Conference, 297-298.

[4] Kosugi, N. et al. "A practical query-by-humming system for a large music database," in ACM Multimedia 2000, Los Angeles, CA, November, 2000.

[5] Feng, Y. Z. Zhuang, Y. T. and Pan, Y. H. "Query similar music by correlation degree," in IEEE PCM2001, 885-890.

[6] Kim, Y. Chai, W. Garcia, R. and Vercoe, B. "Analysis of a contour-based representation for melody," in ISMIR2000, Oct. 2000.

[7] Lu, L. You, H. Zhang, H. J. "A new approach to query by humming in music retrieval," in ICME2001, Tokyo, August 2001.

[8] Foote, J. and Uchihashi, S. "The Beat Spectrum: A New Approach to Rhythm Analysis," in ICME2001, IEEE, Tokyo, August 2001.

[9] Wei Cai, "Melody Retrieval on the Web," Master's thesis, MIT, September 2001.

[10] Lewicki, M. S. "Efficient coding of natural sounds," Nature Neuroscience, Vol. 5, No. 4, pp 356 - 363, April 2002.