Simple But Effective Methods for QBSH at MIREX 2006

J.-S. Roger Jang, Nien-Jung Lee, and Chao-Ling Hsu
Department of Computer Science
National Tsing Hua University, Taiwan
{jang, qmonster, leon}@wayne.cs.nthu.edu.tw

Abstract

This extended abstract describes a submission to the QBSH (Query by Singing/Humming) task of MIREX (Music Information Retrieval Evaluation eXchange) 2006. The methods used for both subtasks 1 and 2 are introduced together with the evaluation results Comments and suggestions for further QBSH task are also addressed in the paper.

Keywords: MIREX, Query by Singing/Humming (QBSH), LS (Linear Scaling), DTW (Dynamic Time Warping).

1. Overview of QBSH Task

The goal of QBSH (Query by Singing/Humming) task at MIREX 2006 is to evaluate MIR systems that take sung or hummed query input from real-world users. QBSH task consists of two subtasks:

- Subtask 1: Known-Item Retrieval
  - Input: 2797 sung/hummed queries of 8 seconds.
  - Test database: 48 ground-truth MIDIs + 2000 Essen Collection MIDI noise files.
  - Evaluation: Mean reciprocal rank (MRR) of the ground truth computed over the top-20 returns.
- Subtask 2: Queries as Variations
  - Input: 2797 sung/hummed queries + 48 ground-truth files of 8 seconds
  - Test database: 48 ground-truth MIDIs + 2000 Essen MIDI noise files + 2797 sung/hummed queries.
  - Evaluation: The precision based on the number of songs within the same ground-truth class of the query calculated from the top-20 returns for each of the 2845 queries.

2. QBSH Corpus

The QBSH corpus provided by Roger Jang [1] consists of recordings of children's songs from students taking the course "Audio Signal Processing and Recognition" over the past 4 years at CS Dept of Tsing Hua Univ., Taiwan. The corpus consists of two parts:

1. MIDI files: 48 monophonic MIDI files of ground truth.
2. WAV files: 2797 singing/humming clips from 118 subjects, with sampling rate of 8 KHz and bit resolution of 8 bits.

For each of the WAV file, the corpus provides another two files distinguished by their file extensions, including PV (files of pitch vectors, derived with a frame size of 256 and zero overlap), and MID (midi files). PV files are pitch vectors labelled manually by the student who recorded the clip. MIDI files were generated from the PV files through a simple note segmentation algorithm. The participants may choose any one of the formats as the input to their systems.

The recording count of each MIDI file is shown in Figure 1.

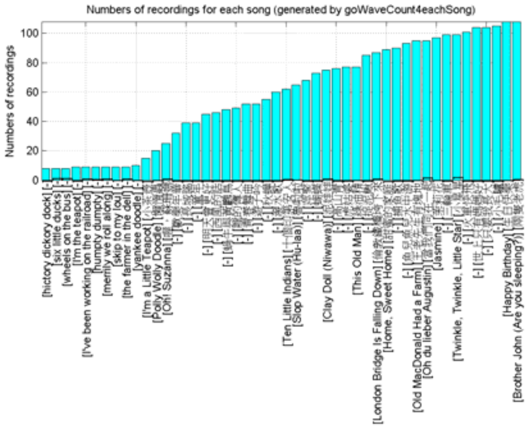

Figure 1. Recording count of each MIDI file.

3. Our Approaches

Since all the query data are available, we have to choose a simple but effective distance measure, which do not run into the potential problem of over fitting/training. Under this guideline, our primary candidates are

- DTW: Dynamic Time Warping [2, 3]
- LS: Linear Scaling [4]
- LS+DTW: LS plus DTW [5]

Then we need to decide which files to be used as the input to our system. Apparently, WAV and PV should be better ones since MID is derived from PV. In order to decide to use WAV or PV, we performed an evaluation similar to subtask 1, where 2000 MIDIs are selected from the Internet as a replacement for Essen Collection. When WAV files are used, we employed a robust pitch tracking algorithm based on dynamic programming to extract the pitch vectors. The result is shown in Figure 2.

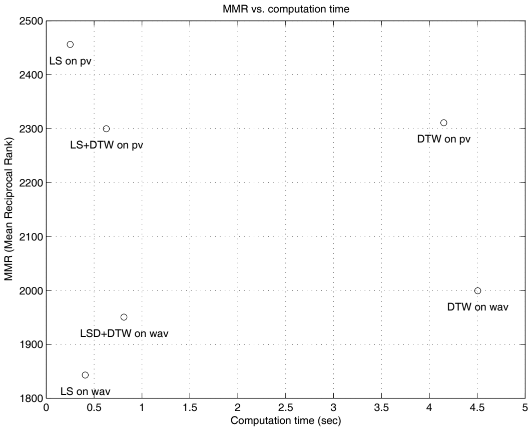

Figure 2. MMR vs. computation time for several methods on PV and WAV files.

Our evaluation demonstrates that PV can always achieve better performance than WAV, as shown in Figure 2, since PV files contain pitch vectors labelled manually. In fact, we still found some mistakes in PV, which should be corrected later in order to make the QBSH corpus more trustworthy. The performance on WAV files is not as good, primarily due to the fact that the WAV files are recorded by 126 subjects at different PCs with different microphone setups, hence it is hard to do both endpoint detection and pitch tracking accurately.

We did not try MID files as the query set since our algorithm is based on pitch vectors (frame-based) instead of music notes.

Since computation time is not really an issue in this task, we used only DTW and LS in our evaluation. Based on the evaluation criteria for both subtasks, we found that LS is the best method for subtask 1 and DTW is the best method for subtask 2.

4. Results

Out simple distance measures do prove to be effective in both subtasks. In subtask 1, an MRR (mean reciprocal rank) of 0.883 is achieved, ranked 3 among 13 participants. For subtask 2, an MP (mean precision) of 0.926 is achieved, ranked 1 among 10 participants. Figure 3 demonstrates the evaluation results for both subtasks.

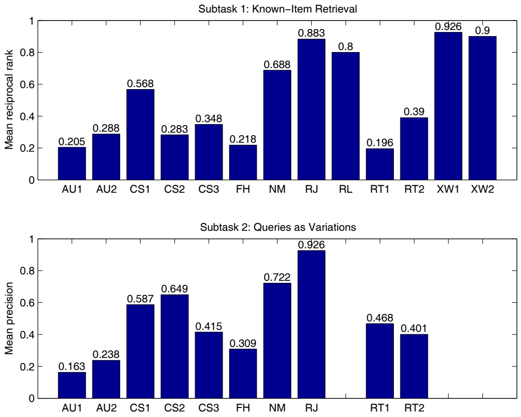

Figure 3. Evaluation results of two subtasks of QBSH

5. Comments on QBSH Task

For a comprehensive evaluation of QBSH in the coming year, we have several comments/suggestions:

Preparation of a test set

Ideally, the test set should not be accessible to any participant beforehand. One way to achieve this is to require every participant to submit a set of recordings to IMIRSEL team to be used as the test set. The test set should be released after the evaluation results are publicized. By following this convention, we should be able to increase our QBSH corpus year by year and new effective methods can be identified accordingly.

More participation

For this year, we have only 13 participants for subtask 1, 10 participants for subtask 2. We should try to encourage more participants since there are much more people working on QBSH.

Variations of QBSH Task

1. Use WAV exclusively as the query input: This is closer to the real-world situation where a QBSH system has to deal with acoustic input to pitch vector conversion using a pitch tracking algorithm.
2. Use MP3 as the test database: This is far more practical then using monophonic MIDIs as the test database. Of course, this is also far more challenging since audio melody extraction is well-known as a tough task in MIREX.

References

[1] J.-S. Roger Jang, "QBSH: A Corpus for Designing QBSH (Query by Singing/Humming) Systems", available at the "QBSH corpus for query by singing/humming" link at the organizer's homepage at http://www.cs.nthu.edu.tw/~jang.

[2] J.-S. Roger Jang and Ming-Yang Gao, "A Query-by-Singing System based on Dynamic Programming", International Workshop on Intelligent Systems Resolutions (the 8th Bellman Continuum), PP. 85-89, Hsinchu, Taiwan, Dec 2000.

[3] J.-S. Roger Jang, Hong-Ru Lee, "Hierarchical Filtering Method for Content-based Music Retrieval via Acoustic Input", The 9th ACM Multimedia Conference, PP. 401-410, Ottawa, Ontario, Canada, September 2001.

[4] J.-S. Roger Jang, Hong-Ru Lee, Ming-Yang Kao, "Content-based Music Retrieval Using Linear Scaling and Branch-and-bound Tree Search", IEEE International Conference on Multimedia and Expo, Waseda University, Tokyo, Japan, August 2001.

[5] Jyh-Shing Roger Jang, Hong-Ru Lee, Jiang-Chuen Chen, and Cheng-Yuan Lin, "Research and Development of an MIR Engine with Multi-modal Interface for Real-world Applications", Journal of the American Society for Information Science and Technology, 2004.

[6] J.-S. Roger Jang, Chao-Ling Hsu, Hong-Ru Lee, "Continuous HMM and Its Enhancement for Singing/Humming Qurey Retrieval", International Symposium on Music Information Retrieval 2005, London, UK, Sept 2005.

[7] J.-S. Roger Jang, Jiang-Chun Chen, Ming-Yang Kao, "MIRACLE: A Music Information Retrieval System with Clustered Computing Engines", 2nd Annual International Symposium on Music Information Retrieval 2001, Indiana University, Bloomington, Indiana, USA, October 2001.