MIREX: Query By Singing/Humming

Jyh-Shing Roger Jang
MIR Lab, CS Dept
Tsing Hua Univ., Taiwan
`jang@cs.nthu.edu.tw`

Nien-Jung Lee
MIR Lab, CS Dept.
Tsing Hua Univ., Taiwan
`khair@wayne.cs.nthu.edu.tw`

Abstract

This extended abstract describes a submission to the MIREX (Music Information Retrieval Evaluation eXchange) in the QBSH (Query by Singing/Humming). The methods for both tasks 1 and 2 are briefly introduced. More detailed analysis of the results achieved in the evaluation will be given in the revised version..

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

1. Feature Set Selection

There are three feature sets available, distinguished by their file extensions, including WAV (original query input), PV (pitch vector files), and MID (midi files). There three sets are available for both the query set (queries from the users) and the database set (candidate songs in the database to be retrieved).

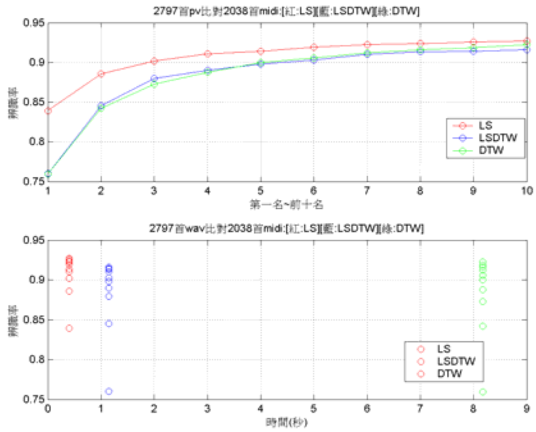
Our evaluation demonstrates that PV can achieve the best performance since they are pitch vectors labelled manually. In fact, we still found many mistakes in PV, which should be corrected later in order to make the QBSH corpus more trustworthy.

The second best is WAV, which is converted to pitch vectors by our dynamic-programming-based pitch tracking. We did try to use PV files as the ground truth to tune the parameters in our pitch tracking algorithm. The recognition rate for pitch tracking seemed satisfactory, but the recognition rate for QBSH is not as good as those achieved by PV directly. We guess that the WAV files are recorded by 126 subjects at different PCs with different microphone setups, 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 instead of music notes.

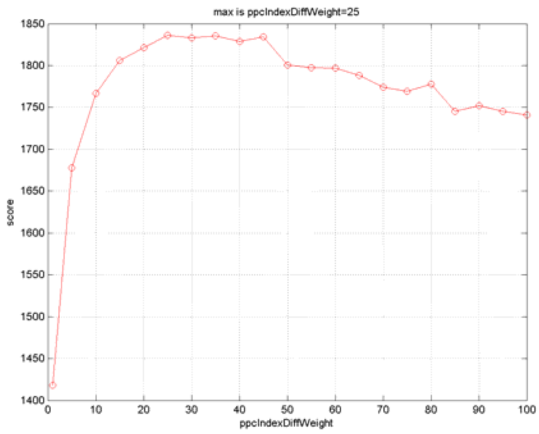
2. Task 1

Task 1 is a traditional problem of query by singing/humming. Here we have tried three methods for this task, including LS (Linear Scaling), DTW (Dynamic Time Warping), and LS+DTW (combination of both). A typical result is shown in the following figure.



In the above figure, PV can achieve a top-10 recognition rate of 92.7%, which is better than those by DTW and LS+DTW. Since the Essen collection is not released, the above evaluation is based on 2000 midi files that we select randomly from our 13000 song collection.

The second plot of the above figure shows the computation time for each method. It should be noted that LS is very efficient in computation, therefore we can perform some parameter tuning (for instance, in pitch tracking) over the whole corpus. We did try this in tuning one of our pitch-tracking parameter, as shown in the following figure:



The above figure demonstrates that when the ppcIndexDiffWeight (one of the parameters in our pitch tracking) is 25, the mean reciprocal rank can achieve a maximum at 1835.9/2797=0.66. However, this figure of merit is not as good as that achieved by using PV files, which is around 0.72.

3. Task 2

Task 2 is a variant retrieval problem that tries to retrieve relevant queries from the query set itself. Again, we have tried the same three methods of LS, DTW, and LS+DTW. The evaluation criterion is the precision based on the returned 20 items. It turned out that DTW has the best performance. However, it should be noted that DTW is computation intensive and it is hard to perform parameter tuning based on the whole corpus using DTW. (More details will be covered in the revised version.)

References

[1] 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.

[2] 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.

[3] J.-S. Roger Jang, Hong-Ru Lee, "I-Ring: A System for Humming Transcription and Chord Generation", IEEE International Conference on Multimedia and Expo, Taipei, Taiwan, June 2004.

[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] J.-S. Roger Jang, Hong-Ru Lee, "Hierarchical Filtering Method for Content-based Music Retrieval via Acoustic Input", The 9th ACM Multimedia Conference (Oral paper, acceptance rate 16%), PP. 401-410, Ottawa, Ontario, Canada, September 2001.

[6] 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.

[7] 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.