MIREX 2012 QBSH Task: YINLONG's Solution

Lei Wang
SHANGHAI YINLONG
INFORMATION
TECHNOLOGY CO., LTD
Leiwang.mir@gmail.com

ABSTRACT

This extended abstract describes my submission to the QBSH (Query by Singing/Humming) task of MIREX (Music Information Retrieval Evaluation eXchange) 2012. The system takes advantage of note-based and frame-based matching methods to improve the accuracy of the Query by Singing/Humming system. First, Earth Mover's Distance (EMD), which is note-based and much faster, is adopted to eliminate most unlikely candidates. Then, Dynamic Time Warping (DTW), which is frame-based and more accurate, is executed on these surviving candidates. Finally, a weighted voting fusion strategy is employed to fusion the result of the two similarity measurement, the final decision is the one with highest scores.

1. INTRODUCTION

There are two sub-tasks in Query by Humming/Singing Evaluation this year. Task 1 deals with the Roger Jang's corpus, the queries are from the beginning of references, 2797 queries from 48 ground truth Midis are used in this task, and the ThinkIT's corpus, the queries are from anywhere of a song, 355 sung queries from 107 Midis are adopted as test set. 2000 Essen Midis are added as noise data in both tasks. As for evaluation, mean reciprocal rank (MRR) of the ground truth is calculated over the top 20 candidates returned by the matchers. Task 2 is variants QBSH evaluation, this is based on Prof. Downie's idea that queries are variants of "ground-truth" midi. In fact, this becomes more important since user-contributed singing/humming is an important part of the song database to be searched, as evidenced by the QBSH search service at www.midomi.com.

2. SYSTEM DESCRIPTION

2.1 Music Score Processing Module

To support humming from anywhere to anywhere of a specified song, phrase segmentation method is adopted to locate entry point of a melody. According to the statistical analysis of a popular song queries set, 98.5% people hum from the beginning of a music phrase. But preprocessing MIDI file and locating the right entry points people likely hum is not an easy task. In our system, we use our hybrid rules including Max Repeated String (MRS) algorithm, pause, long note, theme pattern length to give any potential entry point a weight.

2.2 Acoustic Processing Module

We do not make any constraint on input methods, users could sing lyrics or humming a tune of a specific melody or even could query with meaningless syllables. Since no pitch extraction method is perfect, errors might be introduced during this stage. There are mainly two kinds of pitch errors: 1) double or half of the normal frequency; 2) small pitch fluctuation caused by jitter. To mitigate the impact caused by the two error source, a five-point median filter is adopted to generate smoother pitch sequence.

2.3 Template Matching Module

EMD [3] has robust properties to errors brought about in the front end. Note segmentation mistakes to fragment note into two separate parts. In traditional note-based matching, this action will induce an insertion cost and otherwise a deletion cost from query to template. But in EMD, Such error is trivial because EMD is in a top-down scale of view, the contribution of query notes is approximate equal to the notes of template, no matter how much fragmented or consolidated errors segmentation brings. As for insertion or deletion error, the influence of string matching is two units cost, but EMD is little, either. Moreover, since EMD needs much less computational cost than DTW, to use it as a filtering scheme could also reduce overall searching time greatly. The user's input may be performed faster or slower comparing to the template stored in melody database, thus, the Euclidean distance would produce significant error. To overcome this problem, dynamic time warping (DTW) algorithm [2] is adopted to fill the gap caused by tempo variation between each individual user's acoustic input and music score template. The DTW algorithm is widely used for comparing time series with non-linear distortions in the time axis. This method exhibits good performance in isolate word speech recognition and query by humming. As the finer matching method in our system, DTW algorithm searches the path with the least global distance between query and reference. However, it needs much more computational time than EMD. Although we noticed that although DTW outperforms EMD, the wrong decision made by the two methods would not necessarily be the same. This property suggests that these scores generated through filtering stage could offer complementary infor-

mation which could improve the performance of the final decision. Actually, the combination of different similarity measurement is a way of uncertainty reduction. By doing this, better result could be got than using either DTW or EMD alone.

3. REFERENCES

[1]http://www.music-ir.org/mirex/wiki/2012:Query_by_Singing/Humming

[2] Roger Jang."Hierarchical Filtering Method forContent-based Music Retrieval via Acoustic Input", ACM-MC, 2001.

[3] R.Typke. “Using transportation distances for measuring melodic similarity”. ISMIR. 2003.

[4] Lei Wang. "Improving Searching Speed and Accuracy of Query by Humming System Based on Three Methods: Feature Fusion, Candidates Set Reduction and Multiple Similarity Measurement Rescoring", INTERSPEECH, 2008

[5] Lei Wang. " An Effective and Efficient Method for Query by Humming System Based on Multi-Similarity Measurement Fusion", ICALIP, 2008

[6]Shen Huang “Query By Humming Via Multiscale Transportation Distance In Random Query Occurence Context”, ICME 2008