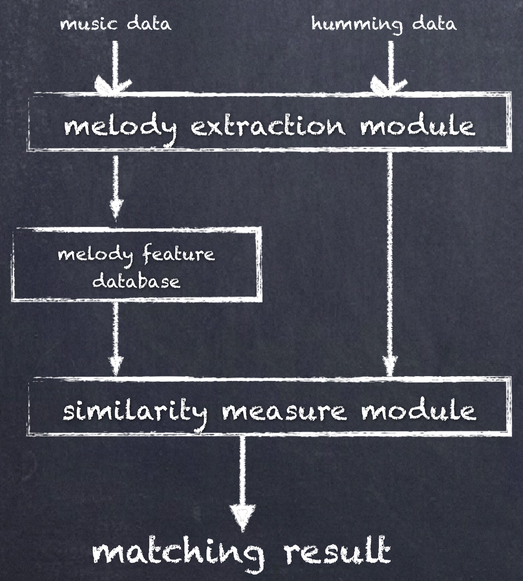
Query-by-Humming System for Polyphonic Audio

Junzhi Du
Department of Electrical and Computer Engineering
dj910906@gmail.com

Abstract

Query-by-Humming(QBH) is one of applications that make users retrieve their objects based on the intrinsic characteristics of music — melody. However, since most related works are aimed to retrieve music from symbolic music data, relatively the works in real world—music information retrieval from polyphonic raw radio, are limited due to its complexity. Our system provides a method to represent melody feature of music by the probability of note occurrence. To measure the similarities and disparities between melody and humming features, we use DP matching method.

Overview of QBH System



MELODY FEATURE EXTRACTION

Step 1: Harmonic Enhancement

extracting predominant pitches

$$E_t^{EP}(k)=\sum_{i=-W}^{W} A(E_t(k)-E_t(k+i)),\ 0\le k<N$$

Step 2: Harmonic Sum

showing periodic pitch

$$F_t(p)=\frac{1}{\left\lfloor \frac{N}{p} \right\rfloor}\sum_{m=1}^{\left\lfloor \frac{N}{p} \right\rfloor} E_t^{EP}(mp)$$

Step 3: Note Strength Calculation

quantizing frequency

$$NS_t(m)=\frac{\int_{L_m}^{U_m} F_t(p)dp}{|U_m-L_m|},\ 0\le m\le M-1$$

Step 4: Note Segmentation

grouping similar pitches

$$SB=\{t \mid \min_t(FE(t)),\ \min(FE(t))<TH\}$$

$$FE(t)=\frac{1}{N}\sum_{k=0}^{N-1} EP_t^2(k)$$

Step 5: Construction of NoteSegment Sequence

obtaining sequence of vectors as possibility values of pitch for segments

$$S_i(m)=\frac{1}{C}\sum_{t=l_s}^{l_e} NS_t(m),\quad 0\le m\le M-1$$

Matching

To overcome the disparities in length between humming query and melody samples and erroneous environment, DP learning is used in the system.

We construct a matrix D of size NR x NQ to show the disparity between music and humming sample.

We could derive a matching path with a specific pitch shift between music and humming sample from D.

After DP matching for each different pitch shift, we can find the matching path with a minimum matching cost.

Windowing method is used to avoid invalid matching, which is far from the ideal diagonal matching path.

$$d_{ps}(r_i,q_j)=\frac{\sqrt{\sum_{m=0}^{M-1}\left[r_i(m)-q_j(m-ps)\right]^2}}{\sqrt{\sum_{m=0}^{M-1} r_i^2(m)}\sqrt{\sum_{m=0}^{M-1} q_j^2(m-ps)}},\ 0\le m,m-ps\le M-1$$

Experiment results

	Top 1	Top 3	Top 5
M01	8.33	25	41.67
M02	8.33	33.33	41.67
M03	8.33	33.33	41.67
M04	8.33	25	41.67
M05	8.33	33.33	41.67
M06	8.33	33.33	41.67

Retrieval accuracy in different experiment configuration

Conclusion and Future work

1. This project is a system of matching humming query with polyphonic data.
2. Melody feature is represented by the probability instead of distinct notes.
3. DP matching is adopted to avoid erroneous notes.
4. Future works could be:
   a. adopting music filters and melody part detection to obtain a better performance.
   b. trying to build a system that deals with complete songs database.