Imperial Journal of Interdisciplinary Research (IJIR)
Vol.2, Issue-1 , 2016
ISSN : 2452-1363 , www.onlinejournal.in

Humming Query Music Search In the Database.

Sanket Chaudhari[1], Mandar Doiphode[2], Akshay Girase[3], Akash Dagade[4] & Prof. S. S. Pawar[5]
[1,2,3,4&5]Sinhgad Academy Of Engineering, Kondhwa, India

Abstract

Recently, the necessity for content-based music retrieval that can return results even if a user does not know information such as the title or singer has increased. Query-by-humming (QBH) systems have been introduced to address this need, as they allow the user to simply hum snatches of the tune to find the right song. Query by humming (QBH) means to search a piece of music by singing or humming. Given melodies hummed by the users, query by humming systems will return an ordered list of songs according to the similarity between humming's and target songs. Query by Singing/Humming (QBSH) is a Music Information Retrieval (MIR) system with small audio excerpt as query. For our work, the input is humming sound which is sound wave and Musical Instrument Digital Interface (MIDI) is used as the reference song in database. In this paper, a system for querying an audio database by humming is described along with a scheme for representing the melodic information in a song as relative pitch changes. We are using Dynamic Time Wrapping Algorithm (DTW) for searching.

Keywords-*Query By Humming, DTW, MIDI, Evolution.*
.

1. Introduction

With the rapid increase in music data on the Internet,
MP3 players, portable media players (PMP), and smart Phones, it is more difficult for the users to find the correct file they want. In addition, if they do not know the information details such as the title and singer's name, more time is needed for searching. To overcome these problems, the query-by-humming (QBH) method has been introduced as a natural interface based on melody that can search for the corresponding music file according to the user's humming. [1]

In this work, audio queries are used to search symbolic (MIDI) targets in a database. Because we use "real" queries from non-musicians, the general quality of queries is low, making retrieval quite challenging. The difficulty of our task and configuration can be seen as a feature: we are far from any ceiling effects, so any significant improvement will show up clearly in the results. Perhaps the most difficult problem for QBH systems is the determination of melodic similarity. Our previous work considered several algorithms for computing melodic similarity. One is based on the fairly well known dynamic programming algorithms for string matching, which had already been applied to the melodic similarity problem. [2, 3] Due to the problems of collecting and annotating large databases, our test database is limited in size. To examine issues of scalability, we estimate performance as a function of the database size. In all cases we have examined, the database precision falls roughly according to $1/\log(x)$ where x is the database size. This is encouraging because this function is flat, meaning performance falls very slowly with increases in database size. [4] Normally, natural sounds are a composition of a fundamental frequency with a set of harmonics. The frequency that the human ear interprets as the pitch of a sound is this fundamental frequency, even if it is absent in the sound. The pitch of natural sounds is important in many contexts. Pitch is the perception of how high or low a musical note sounds, which can be considered as a frequency which corresponds closely to the fundamental frequency or main repetition rate in the signal [5].

It is one of the most important parameters in the voice signal analysis and can be determined by the fundamental frequency of the unit frame. For QBH system, pitch is the key feature of melody. As the humming sound consists of noise, pitch needs to be extracted and in order to get the most significant information. [7]

The main features of this are:

- A melody representation, which combines both pitch and rhythmic information.
- New approximate melody matching algorithms based on the representation.
- A set of automatic transcription techniques customized for the query-by humming system to obtain both pitch and rhythmic information.
- A handy tool to build a melody database from MIDI format.
- A deliverable query-by-humming system including both the server application and the browser application.

In addition to preprocessing the audio queries, we also preprocess the database of MIDI files. Complete
MIDI files include much more than just melodies. In a typical song, the melody is repeated several times.

There is also harmony, drumming, a bass line, and many notes may be performed by more than one instrument, e.g., if a violin and flute may play the same notes, the MIDI file will have separate copies of the notes for the violin and flute. Searching entire MIDI files will take more time and could result in spurious matches to harmonies, bass lines, or even drum patterns. For both speed and precision, we extract melodies from MIDI files and search the melodies. [8]

2. Related Work

In modern days traditional ways of listening to music, and methods for discovering music, are being replaced by personalized ways to hear and learn about music. In other words, it is changing the nature of music dissemination. The rise of audio and video databases necessitate new information retrieval methods tailored to the specific characteristics and needs of these data types. An effective and natural way of querying a musical audio database is by singing/humming the tune of a song. Even though a variety of QBSH [9,] techniques have been explored, there has been relatively less work on analysis of QBSH system through query excerption.

Many QBSH techniques represent song or piece of song as point sequences [4], Hidden Markov Models (HMMs) [7], Modified Discrete Cosine Transform (MDCT) coefficients and peak energy. Few use loudness model of human hearing perception along with local minimum function to recognize onsets in music database and query. In other works music is treated as a time Series and employed a time series matching approach because of its effectiveness for QBSH in Terms of robustness against note errors. In, authors retrieve the melody of the lead vocals from music databases, using information about the spatial arrangement of voices and instruments in the stereo mix. The retrieved time series signals are approximated into symbolic strings which reveal higher-level context patterns. The method elaborated in builds an index of music segments by finding pitch vectors from a database of music segments. In another work, a method for showing the melodic information In a song as relative pitch changes is projected. Further, the work is based on extracting the pitch out of monophonic singing or humming, and later segmenting and quantizing the information into a melody composed of discrete notes. Some [4, 6, 8] of them concentrate on developing QBSH systems using mammoth music collections. The objective is to construct a dependable and well-organized large-scale system that collects thousands of melodies and responds in seconds. The query melody searching and alignment is done using skeletons of the melody [6].

To complement the work proposed a technique using melody matching model based on the genetic algorithm and improving the ranking result by Local Sensitive Hashing (LSH) algorithm. Authors employed double dynamic programming algorithm for feature similarity matching.

3. Music Search Method.

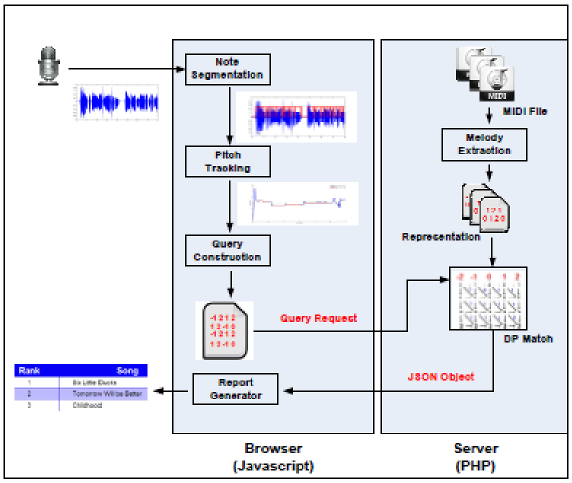

Figure 1. Basic Architecture

This is the primary step of classification process, where the document, like .pdf, .doc, .html, is taken as an input and it is scanned well and forwarded to the pre-processing unit.

3.1 Browser Side Architecture

Note Segmentation: The purpose of note segmentation is to identify each note's onset and offset boundaries within the acoustic signal. In order to allow segmentation based on the signal's amplitude, we ask the user to sing using the syllables like da, thus separating notes by the short drop in amplitude caused by the stop consonant.

Pitch Tracking: This component is to estimate the pitch or fundamental frequency of a periodic or virtually periodic signal, usually a digital recording of speech or a musical note or tone. This can be done in the time domain or the frequency domain or both the two domains.

Query Construction: The note information will be transferred to a string representation and send to the server as a HTTP GET request.
http://localhost/query.php?seq='-121012-2'
Report Generator: Parse the JSON object returned by the server side and generate a rank list report to the user.

3.2 Server Side Architecture

Music Database: This includes the source data. The source data are the original music corpora, from which we extract melodies and generate the data representation. The source data are in MIDI files currently (but it can be Extended to handle other music format).

Melody Extractor: This extracts the melody information from MIDI file and transfer to a string sequence to represent the note information.

DB Matcher: This receives the query from browser side, uses dynamic programming to match it with the melodies in the melody description objects, and returns a rank list for matching songs to the user as a JSON

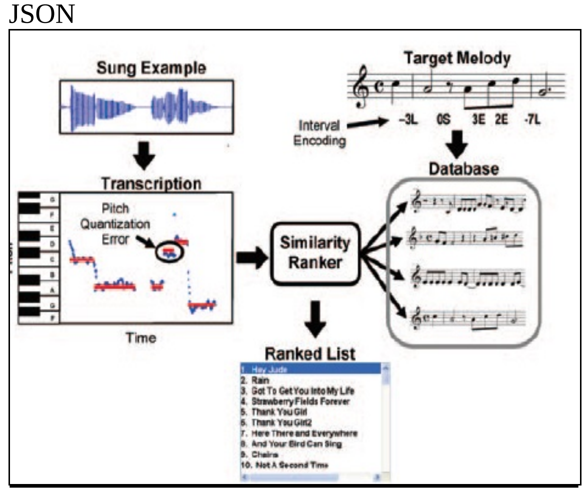

Figure 2. Purposed System

3.3 Pre-processing

MP3 songs contain convoluted melody information and even noise. Thus pre-processing is applied on the MP3 songs database to extract information needed by the system. Most of the MP3 songs possess 44.1 KHz sample rate and dual-channel data, but for melody representation, such high quality signal is not necessary. Also it will make further processing time consuming and inefficient. In fact, even in very low sample rate, melody of the songs can be identified. The International Journal of Multimedia & Its Applications Therefore, MP3 songs are decoded into wave streams, down sampled to 8 KHz and converted to mono channel.

3.4 Vocal and Non Vocal Separation

In music, human vocal part always plays an important role in representing melody rather than its background music, it is desired to segregate both. Furthermore, music researchers have shown that the vocal and non-vocal separation exploits the spatial arrangement of instruments and voices in the stereo mix and could be described as inverse karaoke effect. Most karaoke machines adopt center pan removal technique to remove the lead voice from a song [11]. One stereo channel is Inverted and mixed with the other one into a mono signal. The lead voice and solo instruments are generally centered in the stereo mix whereas the majority instruments and backing vocals are out of center. Above mentioned transformation is used to remove the lead voice from music. We intend to invert this effect, so that the pre-processed song yields a high portion of the lead voice, while most other instruments are isolated and removed. For vocal and non-vocal separation audio editor's voice extractor option with center filtering technique is employed.

3.5 Feature Selection and Extraction

Extracting significant feature vectors from an audio signal is a major task to produce a better retrieval performance. In this work, Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Coefficients (LPC) and Linear Predictive Cepstral Coefficients (LPCC) features are preferred because they are promising in terms of discrimination and robustness. For these features frame size and hop size that is the span between the starting times of two succeeding frames are empirically determined for the retrieval. Most feature extraction techniques produce multidimensional feature vector for every frame of audio.

4. Conclusion

In this project, we build a web-based query-by-humming system, use pitch tracking and dynamic programming matching method. For future work, we need to test our system on different classes of music (i.e. pop music, country music). We also can try other matching methods like DTW (Dynamic Time Warping), HMM (Hidden Markov Model) and compare the performance.

5. References

[1] 1. JH McClellan, RW Schafer, MA Yoder, Signal Processing First (Pearson Prentice Hall, 2003)

[2] W. Hewlett and E. Selfridge-Field, "Melodic Similarity: Concepts, Procedures, and Applications," in Computing in Musicology, vol. 11. Cambridge: MIT Press, 1998.

[3] R. B. Dannenberg, "An On-Line Algorithm for Real-Time Accompaniment," in Proceedings of the 1984 International Computer Music Conference, Paris, San Francisco: International Computer Music Association, 1985, pp. 193-198.

[4] A Comparative Evaluation of Search Techniques for Query-by-Humming Using the MUSART Testbed Roger B. Dannenberg Carnegie Mellon University, rbd@cs.cmu.edu William P. Birmingham Grove City College Bryan Pardo Northwestern University Ning Hu Google, Inc. Colin Meek Microsoft Corporation

[5] Mcnab R., Smith L., and Witten I., “Signal Processing for Melody Transcription,” in Proceedings of the 19th Australasian Computer Science Conference, Australia, pp. 301-307, 1996.

[6] Jun B., Rho S., and Hwang E., “An Efficient Voice Transcription Scheme for Music Retrieval,” in Proceedings of International Conference on Multimedia and Ubiquitous Engineering, Seoul, pp. 366-371, 2007.

[7] Jing Q., Wang X., Zhou M., and Liu X., “A Novel MIR Approach Based on Dynamic Thresholds Segmentation and Weighted Synthesis Matching,” in Proceedings of IET Conference on Wireless, Mobile and Sensor Networks, Shanghai, pp. 1017-1020, 2007.

[8] Kim H. and Sikora T., “Audio Spectrum Projection Based on Several Basis Decomposition Algorithms Applied to General Sound Recognition and Audio Segmentation,” in Proceedings of the 13th European Signal Processing Conference, Austria, pp. 1047-1050, 2004.