Proposal for OUR Project 2016-17

Supervisor: Dr. N Sukumar - Director, Center for Informatics and Professor, Department of Chemistry - n.sukumar@snu.edu.in
Student: Ganesh S - BTech Computer Science and Engineering - gs401@snu.edu.in

Title of the proposal:
Designing a robust and scalable Query By Humming system by using a sequence of wavelet-transform coefficients and Levenshtein edit distance.

Scientific background and context:
How do you find the name of a piece of music that's stuck in your mind? For most humans, humming is a natural way to represent music. It is how they query someone who knows many songs. Can we develop an algorithm that does this? Receive user's tune, process it, and retrieve the correct song. The **query-by-singing/humming problem** is to retrieve music from a given melodic or rhythmic input (in audio or symbolic format), extract features, compare them to those in the database and retrieve the correct song.

A system which solves this problem needs to be robust to poor humming (wrong pitch, wrong note duration, wrong key) as well as noise.

Significance and Innovation:
Less than ten query-by-singing/humming problem (QBSH) systems have been published till date and availability of rigorous evaluation results for these is scarce. The most popular one is a commercial software called SoundHound that matches users' hummed queries against a proprietary database of hummed songs.

Discovery of better features that uniquely identify songs may improve performance in other tasks within the music information retrieval research such as audio melody extraction and similarity.

Specific goal:
The specific aim of this project is to is find a novel and better-performing solution to a problem in the field of music informatics: query by singing/humming to retrieve music.

The most common techniques [1] used by existing solutions are hidden Markov models, melodic contour (use Dynamic Time Warping to align the pitch contour of the query with that of the target), n-gram, note interval matching (treating melodies as strings and using dynamic programming to align two strings). Similar techniques have also been applied in the fields of proteomics, bioinformatics and cheminformatics.

We will explore the use of machine learning techniques such as the k-means clustering algorithm to solve this problem.

We propose to use publicly available data sources, such as *Roger Jang's MIR-QBSH* corpus [2], which consists of 4431 queries along with 48 ground-truth MIDI files. Such datasets are used in *MIREX (Music Information Retrieval Evaluation eXchange)* : an annual competition for solutions to this and other music information retrieval problems [3]. These contain popular songs and user-hummed queries. For specific genres such as carnatic music, there are open datasets such as the *Carnatic Music Kriti Audio Archive* [4].

Broad goal:

An investigation of Query-by-Singing/Humming problem and study of dependant/related problems such as music feature extraction (melody-harmony-tonality, onset detection and segmentation, rhythm), music similarity (self-similarity, global similarity) and music classification (instrument, genre, mood, culture) [5].

Summary of Research Plan and Milestones:
Since rigorous evaluation data for existing systems is not easily available, the first task would be to do this evaluation by reproducing existing publications (one implementation for each of the four most common applications). This should take not more than a month.

Next, we plan to evaluate how various types of preprocessing of input data to obtain descriptors (only extract melody, extract melody and beat information, sampling or smoothing methods, etc.), affect performance. This second milestone should also take one month.

After this, we will implement a novel technique for a specific genre, *raga*-based Carnatic music (which might not lend itself well to the usual up/down/same melodic contour [6]). We will use the discrete wavelet transform on the recorded audio signal (both database songs and the query) to get a small number of wavelet coefficients, thereby eliminating arbitrariness. We will generate a time series of the ratios of successive wavelet coefficients (in modulo 2 arithmetic) to represent the melody. This time series would be invariant to shifts in tonic or key. A distance metric such as Levenshtein edit distance would be the most natural way to quantify similarity of the query sequence to the different sequences in the database. This distance metric, which is computed using a dynamic programming algorithm, counts the minimum number of removal, insertion or substitution operations required to transform one string/sequence into the other. It is commonly used in sequence alignment algorithms (like BLAST - Basic Local Alignment Search Tool) in bioinformatics.

	Before Aug	Aug-16	Sep-16	Oct-16	Nov-16	Dec-16	Jan-17	Feb-17	Mar-17	After Mar-17
Literature survey for this particular problem	Already done									
Implementation of existing techniques and tests against single dataset										

Comparing performance by varying preprocessing and combinations of descriptors										
Implementation of wavelet-coefficient-ratios idea and testing against Carnatic music dataset										
Study of ML techniques used for similar problems in music, text, video retrieval										
Implementation and testing of different solutions										
Report writing										
Submission to MIREX evaluation competition										Not part of OUR

Resources Required

We request permission for the student to use a computer in the Informatics lab. It is estimated that a total of four hours of computing time on 32 nodes of the Magus High Performance Computing cluster would be required. The reason for requesting HPC time is because the MIREX challenge evaluations also take place on nodes with significant computational power. This would also give the student practical experience with development of parallel code for a real life application.

References

[1] Eugene Weinstein. Query by Humming: A Survey. NYU Computer Science. 2005. https://cs.nyu.edu/~eugenew/publications/humming-summary.pdf

[2] Roger Jang. Query By Singing/Humming - Music Information Retrieval corpus. 2006.

[3] MIREX. Music Information Retrieval Evaluation Exchange. 2015.

[4] Shivkumar Kalyanaraman. Carnatic Music Kriti Audio Archive. 2008

[5] Schedl, Gomez, Urbano. Music Information Retrieval: Recent Developments and Applications. 2014.

[6] N Sukumar, T V Gopalakrishnan. The Algebraic Structure of Carnatic Indian Music.