MUSICAL SCALE DEDUCTION USING MFCC

KAUSTAV SANYAL

Abstract: This paper is the preliminary phase of the complete undergoing project of Vedic Swara recognition system, as this deduces the musical scale of a chant. Before deducing the swaras for a Vedic chant we must know in which scale the complete chant is being performed. Without knowing the scale it will be difficult to deduce the accuracy of the chant. We have used a popular technique named MFCC (Mel-Frequency Cepstral Coefficient) for this purpose. In this paper we will analyze the musical scale for the chant and obtain the result for further use. This is just the first phase of the pipelining of the complete Vedic Swara Recognition System.

Keywords: Musical scale, Vedic Swara, Mel, Cepstral.

Introduction: One of the oldest languages of this world, Sanskrit has one of the richest vocabulary and literature collections in world. This language can be divided broadly into two divisions namely, the Vedic Sanskrit and the Laukik Sanskrit. The alphabets (Varnas) of Sanskrit language represent the most unambiguous and accurate representation of speech. But with all its ancillary parts, the Vedic Sanskrit consists of a very specific and important grammatical term, Swaras. The Swaras are one of the most difficult grammatical parts of Vedic Sanskrit, that being a very scientific way to pronounce the words enhances human intellect and knowledge towards the understanding of Vedic Suktas. It is believed that the most accurate pronunciation of Vedic Suktas automatically generate its meaning in human minds. But the Swara system is very difficult to recite and each mistake in uttering the swaras may lead to complete different meaning of the recited Sukta. The Vedic Sanskrit, being world's one of the most complex language deviates a lot from the Laukik Sanskrit in its vocabulary and sentence structures. The construction of the Vedic suktas applies a completely independent grammar for its sentence constructions and Swara composition. It mainly includes the rules from the Panini Ashtadhyayi. But all these are possible only if the person chanting maintains a specific musical scale during the chant. Otherwise the deviation from notes may lead to an inaccurate chant. Here we will recognize the musical scale of the chant using the MFCC technique.

Background Theory: 2.1 Musical Scales: In music theory, a scale is any set of musical notes ordered by fundamental frequency or pitch. A scale ordered by increasing pitch is an ascending scale, and a scale ordered by decreasing pitch is a descending scale. Some scales contain different pitches when ascending than when descending. For example, the Melodic minor scale. Often, especially in the context of the common practice period, most or all of the melody and harmony of a musical work is built using the notes of a single scale, which can be conveniently represented on a sta_ with a standard key signature. Scales are typically listed from low to high. Most scales are octave-repeating, meaning their pattern of notes is the same in every octave (the Bohlen~nPierce scale is one exception). An octave-repeating scale can be represented as a circular arrangement of pitch classes, ordered by increasing (or decreasing) pitch class. For instance, the increasing C major scale is C-D-E-F-G-A-B-[C], with the bracket indicating that the last note is an octave higher than the _rst note, and the decreasing C major scale is C-B-A-G-F-E-D-[C], with the bracket indicating an octave lower than the _rst note in the scale. Now this thing is similar to the Indian Saptak Concept as it rounds from Sa to Sa. Scales may be described according to the intervals they contain: for example: diatonic, chromatic, whole tone or by the number of di_erent pitch classes they contain:

1. Octatonic (8 notes per octave): used in jazz and modern classical music.
2. Heptatonic (7 notes per octave): the most common modern Western scale.
3. Hexatonic (6 notes per octave): common in Western folk music.
4. Pentatonic (5 notes per octave): the anhemitonic form (lacking semitones) is common in folk music, especially in oriental music; also known as the "black note" scale.
5. Tetratonic (4 notes), tritonic (3 notes), and ditonic (2 notes): generally limited to prehistoric ("primitive") music.
6. Monotonic (1 note): limited use in liturgy, and for e_ect in modern art music.

The number of the notes that make up a scale as well as the quality of the intervals between successive notes of the scale help to give the music of a culture area its peculiar sound quality.The pitch distances or intervals among the notes of a scale tell us more about the sound of the music than does the mere number of tones.

2.2 MFCC: In this project the most important thing is to extract the feature from the audio signal. The audio signal feature extraction in a categorization problem is about reducing the dimensionality of the input-vector while maintaining the discriminating power of the signal. In the feature extraction or signal processing stage, the acoustic waveform is sampled into frames (usually of 10, 15, or 20 milliseconds) which are transformed into spectral features. Mel: Human hearing, however, is not equally sensitive at all frequency bands. It is less sensitive at higher frequencies, roughly above 1000 Hertz. It turns out that modeling this property of human hearing during feature extraction improves speech recognition performance. The form of the model used in MFCCs is to warp the frequencies output by the DFT onto the mel scale. A mel (Stevens et al., 1937; Stevens and Volkmann, 1940) is a unit of pitch deed so that pairs of sounds which are perceptually equidistant in pitch are separated by an equal number of mels. The mapping between frequency in Hertz and the mel scale is linear below 1000 Hz and the logarithmic above 1000 Hz.

Cepstral: One way to think about the cepstrum is as a useful way of separating the source and filter. The speech waveform is created when a glottal source waveform of a particular fundamental frequency is passed through the vocal tract, which because of its shape has a particular filtering characteristic. But many characteristics of the glottal source (its fundamental frequency, the details of the glottal pulse, etc) are not important for distinguishing different phones. Instead, the most useful information for phone detection is the filter, i.e. the exact position of the vocal tract. If we knew the shape of the vocal tract, we would know which phone was being produced. This suggests that useful features for phone detection would and a way to deconvolve (separate) the source and filter and show us only the vocal tract filter. The cepstrum is more formally defined as the inverse DFT of the log magnitude of the DFT of a signal.

Materials and Methods: Used softwares to make the training corpus: FLStudio: it is used to take the midi vamp input for a specific scale.

Adobe Audition CC2014: it has been used to record the sample training sets and make the training corpus.

The steps are as follows:

1. Open a new audio file with settings of channel: mono, bit depth: 8 bits
2. Open a new multitrack composition
3. Place the required scale vamp
4. Record accordingly
5. Save as .wav file

Matlab: for the main coding of the program which has two phases namely training and testing. Two separate program codes have been written for the steps.

Training: To start the classification process soundtrain.m file is executed. Matlab command window prompt for input the directory name where training file has stored. After giving input it read every file in the folder by calling a function loaddata.m. Then it call another function mfcc, mfcc function call another function "melcepst.m" which is a standard function for calculating cepstral coefficient and this program has been collected from a matlab tool box voice box. After completion mfcc store cepstral coefficient in a matlab workspace variable cepstral.m and it also store name of the corresponding file. This is the end of training phase.

Testing: After completion of training phase an unknown file is taken soundtest.m file is executed. This file prompts for input sound file. After giving its proper location sound file is read. And then soundtest.m call function mfcc.m. Which calculate mel frequency cepstral coefficient by calling melcepst.m as before. Then cepstral coefficient is passed through a function distmeasure.m along with train cepstral coefficients which were stored in cepstral.mat. Distmeasure.m call function disteusq.m which is a voice box matlab function. This function calculates Euclidean or mahalanbis distance between test file and each of train file. After getting minimum distance it returns an index, this index return name of the file that is store on matlab workspace variable name.mat. By processing the filename we get corresponding match.

Result and Discussion: In this step it is seen how much samples are correctly classified and thus the performance of the system is evaluated. In the testing phase, the approach discussed above was applied to two verses in two different scales namely E and Gm. There are no common sample in training set and testing set. The hit rate is percentage of times the verse was correctly identified. It was found that the proposed approach gave sufficiently good results for all of the verses. Percentage of misclassification was significantly low. Overall the result was satisfactorily good giving the total percentage of accuracy over 97.

The purpose of this paper was to recognize the musical scale of a specific chant. The results show that the features used here can give good classification performance. Though the database of collection of sound is not so large but the variation of sound is so diverse that the above performance makes a satisfactory result.

References:

1. Ibrahim Patel, Dr. Y.Srinivas Rao, Speech Recognition Using HMM with MFCC- An Analysis Using Frequency Spectral Decomposition Technique, Signal & Image Processing: An International Journal(SIPIJ) Vol.1, No. 2, December 2010.
2. Garima Vyas, Malay Kishore Dutta, Hicham Atassi, Radim Burget, "Detection of chorus from an audio clip using dynamic time warping algorithm", Engineering and Computational Sciences (RAECS) 2014 Recent Advances in, pp. 1-6, 2014
3. T. R. Jayanthi Kumari, H. S. Jayanna, "Comparison of LPCC and MFCC features and GMM and GMM-UBM modeling for limited data speaker verification", Computational Intelligence and Computing Research (ICCIC) 2014 IEEE International Conference on, pp. 1-6, 2014
4. Gill KZ, Purves D (2009) A Biological Rationale for Musical Scales. PLoS ONE 4(12): e8144. doi:10.1371/journal.pone.0008144

Kaustav Sanyal
Bankura Unnayani Institute of Engineering.