RECOGNITION OF DASTGAH AND MAQAM FOR PERSIAN MUSIC WITH DETECTING SKELETAL MELODIC MODELS

Nima Darabi, Navid Hamed Azimi, Dr. Hassan Nojumi
Faculty of mathematical sciences, Sharif University of technology

Abstract: In this paper we investigate the use of pattern recognition techniques for recognizing two features of Persian music, dastgah and maqam, by their relation to musical scale and mode. We define some statistical measures that characterize the melodic pattern of a musical signal. We then use a standard perceptron for pattern recognition. This technique could be generalized for recognizing such features in music of other nations and also used in applications like content-based search and indexing of musical databases.

Keywords: musical mode, musical scale, Persian music, dastgah, maqam, content based musical retrieval

Introduction

The tradition of Persian art music embodies twelve modal systems, known as dastgah. Each dastgah represents a complex of skeletal melodic models, called maqam, on the basis of which a performer produces extemporized pieces.

Automatic pattern recognition and clustering techniques have been successfully applied to various kinds of objects. In this paper we define a measure of similarity between a given musical piece and a skeletal melodic model. We use this measure to investigate the use of pattern recognition techniques for revealing dastgah and skeletal melodic model of a given musical piece. The technique introduced in this paper can be extended further to define a similarity measure of any given musical piece (not necessary Persian) and a melodic pattern and further can be used to build a system that enables user to give a musical piece as input and search for the music's that are similar in melodic pattern and in profiling the users based on their previous downloads form a musical database.

Our aim is to develop a system able to recognize dastgah and maqam for any given Persian music piece. For simplicity we assume that the input to this system is in the form of musical signal. This assumption can be made without lose of generality since any other kind of digital music could be easily converted to signal form. We start by extracting proper data that characterize intervallic structure and melodic patterns of the given piece. We use this data to define a similarity measure between two pieces of music. We use this measure and other information's extracted form input to build a proper feature vector for pattern recognition.

In this research we use the previous work of Professor Farhat in analyses of the intervallic structure, melodic patterns, modulations, and improvisations within each dastgah in Persian music.

Basic Definitions

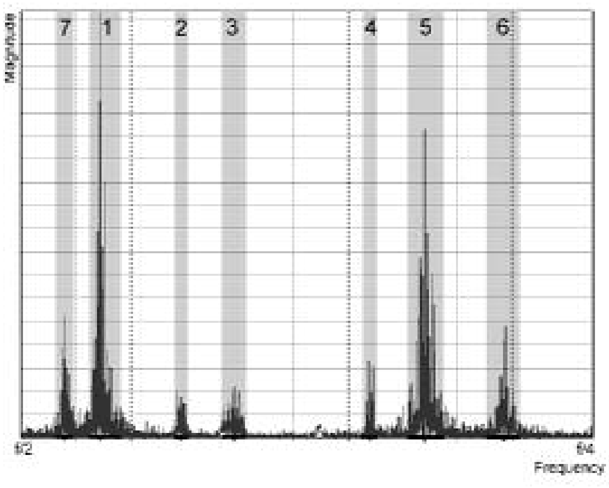
In music, an **Octave** is the interval between one musical note and another with half or doubled frequency. The **Tonic** is the pitch upon which all other pitches of a piece are hierarchically centered. It is usually the pitch that the final sentences of piece end in it. The music **Scale** which will be referred as scale can be defined as a series of notes differing in pitch according to a specific scheme usually within an octave. Musical **Mode** which be referred as mode is any of certain fixed arrangements of the diatonic tones of an octave, an example is the major and minor scales of classical music. Usually only specific frequencies is used in a musical piece among all possibilities. An octave is a frequency range containing all that possible notes. An accurate definition of mode of a certain musical piece can be explained by an array which represent the degrees between consecutive tones of the piece, starting from the tonic, e.g. we can describe a sample normalized form of the Persian mode Shur as (6 7 9 9 4 9 9). It means that If we divide an octave in to 53 equal interval the musical distance between the tonic and the fourth note of the scale Shur will be (6+7+9)/53 octave.

In similar definitions in Persian music we are involved with two basic definitions **Dastgah**, and **Maqam**. Dastgah revolve around unspecified central nuclear melodies which the individual musician comes to know through experience and absorption. All of these melodies are from the same scale and each piece of that scale will be classified in that Dastgah, if can be recognized as a Persian music. Therefore we can use a scale recognition algorithm for revealing the dastgah of a given Persian music. Moreover, each dastgah can be divided into more detailed categories named maqam. There are no clear boundaries between two maqams; besides, different parameters are used to describe two distinct maqams. Some Maqams represent the concept of a mode and some of them are more detailed and have specific melodic characteristics, named **Skeletal Music Model**.

Musical Features

In this part we extract the feature vector necessary for our pattern recognition algorithm. In general every music file contains data from two different domains, frequency and time. One would think that any given feature encodes some information in frequency domain, time domain or both. Most of the features used in previous works on music content based search and classification are in fact encoding information's in both domains [2]. Our first step is to encode the information in the frequency domain into a proper set of features

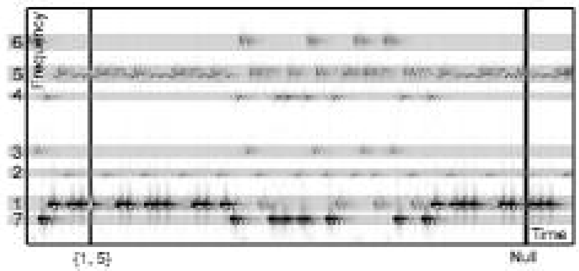
We start by performing fast Fourier transform on our input signal to get frequency domain content of input signal. This process without the mapping part is used in wave to midi converter program frequently. In a musical piece a same note can be played in different octaves. In this research only the notes have been played are important not the octave they have played within, therefore we map the frequency information on a certain octave. A frequency range of F and 2F can be used for any given F as an octave (In realty this frequency range could have notes from two octaves of a real musical instrument). In our program the frequency ranges of f/4 and f/2 as destination octave where f is the sample rate of our input signal. Because of logarithmic nature of the octave, the next octaves would be f/4 to f8, f/8 to f/16, and so on. Because of the insufficient resolution of samples in lower octaves, we only map between 3 to 9 octaves due to sampling frequency and the content of lower frequencies of the signal. Then we calculate algebraic addition of all data to obtain a diagram as figure 1.



As you can see there are certain distinguishable peaks on figure 2. These peaks are actually the diatonic notes used in piece and they mark the frequency intervals of the piece. The non diatonic notes would have little or no effect on this diagram due to infrequent use. With a peak detection algorithm [3] we extract a range for each peak and then calculate the median of the distribution of magnitude over each range as the place of the peak. The range is an interval which contains noticeable energy of the peak an array of these medians can be used as the first entrances of the feature vector and name it mode_array. We then perform some additional process described later to find the tonic note and we update this array so that it represent the number of scale degrees between consecutive tones of the mode, starting from the tonic. We represent each range with a number. This number, representing a note within octave, is index of the array of the medians after we finds the tonic note and shifted the array.

While the feature vector we explained above is enough for mode detection and therefore recognizing dastgah of the input, recognizing skeletal melodic model needs some other information like melodic pattern of the music. This data certainly cannot be found in frequency domain so we need some extra information on time domain.

For extracting melodic patterns and other information's in time domain, we split the input file into several frames and perform a fast Fourier transform on each frame. We then map every frame to a certain octave as before. A visualization of this short-time Fourier transform is shown in figure 2.



The peaks in each frame are the notes played during the frame. With some improvements this technique was used to extract the symbolic representation of a musical signal. We then detect the peaks in each frame. The number of this peaks might vary form zero to number of notes due to different notes played in the frame but it hardly exceed 5. We extract the notes corresponding to the ranges contain the peaks. We ignore every peak not within the ranges .we put the note with the highest energy in its peak in an array named melodic sequence. We then put all the notes of this frame in a set. We then put this information in a sequence of sets a sample of this sequence can be find in figure 2.

This is our initial sequence. We then perform following transformation to omit redundant data. We omit every note stored in an entry if it exists in last entry of the sequence. We also omit every entry that contains no notes. We repeat the whole process until there would be no change in the sequence after an iteration. This sequence contains the information of melodic pattern of our input. We call this sequence Melodic_Pattern.

Beside these mode_array and melodic_pattern we compute some other statistical features as well. These features are:

- Smoothness: This feature is autocorrelation of melodic_sequence and is a measure of smoothness in changing consecutive notes' pitch.
- Deviation: This feature is the variance of melodic_sequence and is a measure of deviation of notes around tonic.
- Tonic: This is the number of tonic note.
- Shamed: This is the number of the note has been played more frequently than the others. This note can be the tonic note or not.
- F_array: This array contains the number each note has been appeared in music sequence.

Detecting scale

We rewrote the definitions for Persian modes obtained from project Scalar [11] with respect to Professor Far hat's research [10] and some other intuitive experiences. Here are some samples of our database represents by dividing an octave into 60 intervals:

(6 9 10 10 5 10 10) Shur, Dashti, Bayate Kord, Abu Ata
(10 10 5 10 10 7 8) Bayate Tork
(10 6 9 10 10 5 10) Afshari
(10 5 10 10 6 9 10) Nava
(6 14 5 10 5 10 10) Homayun
(10 5 10 10 6 14 5) Esfahan
(10 5 10 10 6 11 8) Mokhalefe Segah, Esfahane Ghadim
(10 5 10 10 6 14 5) Bidad
(10 6 14 5 10 8 7) Ist -e- Dovom –e- Homayun
(7 10 7 8 10 10 8) Segah
(7 13 5 10 6 14 5) Chahargah

Now, we just need mode_array to detect the scale and therefore dastgah of a given piece. We must compare the mode_array with any record of our database as reference. First, one can normalize these two arrays so that the summary of elements of each one become 1 and then compute the quadratic sum of differences between any element of mode_array and its closest element in reference array. Divided into the size of mode_array, it will be a measure of difference between these two arrays.

So far we were able to identify dastgah of our input. The next step is to recognize maqams for a given piece. Note that there are some maqams that have their own scales and therefore already recognized such as Ist -e- Dovom –e- Homayun.

Recognizing of skeletal melodic models

There is no clear, mathematical definition of skeletal melodic model refer to as model in rest of the paper. In musicians world, recognizing of the models is usually done with identifying tonic node and a since that musician get from the general shape of the melody such as a sense of up rising, down rising or smoothness. Since we detect the scale and dastgah before with complete accuracy, we can limit our search space to the models within a certain dastgah. Melodic patterns sequence is encoding information about the melodic pattern but it is so large that cannot be used as input for classic pattern recognition techniques such as Euclidean distance or neural networks. To overcome this difficulty we use a new parameter Sim which calculated using a modified version of approximate string matching algorithm

Our algorithm is as follows. We call the search piece sink and the file we are searching in it source as it is common in string matching algorithms. This algorithm can be used for content based music search but this is not the focus of this paper.

```
Sim=0
For i=0 to size of melodic_pattern sequence for source piece do
Sum=0
For j=0 to size of melodic_pattern of sink music do
        Find the minimum of (a – b)²
        where a is form the set of
        entry melodic_pattern [i+j]
        of source file and b is form
        the set of entry
        melodic_pattern [j] of sink
        and add it to sum
Sum=sum/size of melodic_pattern of sink
If sum is smaller than β then Sim = Sim + (1 – 1/(1+sum))
Sim=Sim/Size of source melodic_pattern
```

This algorithm gives us a number Sim which can be thought as a measure of the times part of sink piece is played in source file. By changing β we can change how much of the melody in the sink is actually played in source. In the rest of this paper we assume that β is 0, hoping that if two totally different tones are to be matched, Sum would be large enough to have almost zero effect on Sim.

We store short melodies of each musical model in our model database. We run the algorithm above for the given piece as source and each of the melodies for a certain model once. We add all the Sims from each run of the algorithm together and call it Model_Similarity. It is the last number in our feature vector. This number is actually a measure of how much a given music sequence is like the famous melodies of the musical model. This feature is one of our important measures in our classification process.

Now we can construct the feature vector needed for mode detection. Our feature vector contains Smoothness, Tonic, Shahed, F_Array and Model_Simlarity of the input and each model of the detected dastgah. We use a neural network with hidden layer with back propagation learning algorithm. A different network for each dastgah is trained and the size of hidden layer is set to one fourth of the size of the models within the correspondent dastgah.

Results

For testing our approach we focus on a dastgah named homayon. This dastgah contains 12 maqams which is considerable compare to the others and therefore is suitable for our test. We store 30 short melodies for all the maqams of our database. We collect 70 different pieces from different known maqams in this dastgah. The preliminary process recognizes the right dastgah (Homayon) for 64 pieces in 11 different maqams. For the second part, detecting maqam, the network is trained and tested with the leaving one out technique; the network was trained using 55 pieces and tested on the remaining 9 pieces. 5 tests were run each using defend training and testing database. An average error rate of 28.2% is obtained. It is a dependable result with respect to complexity of the job and similarity of maqams of Homayon.

References

[1] A. Rauber and M. Fruhwirth. "Automatically analyzing and organizing music archives", pages 4–8. 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001). Springer, Darmstadt, Sep 2001.

[2] Ponce de Leon, P.J., I˜nesta, J.M. "Feature-driven recognition of music styles. In 1st: Iberian Conference on Pattern Recognition and Image Analysis", LNCS, 2652. (2003) 773–781

[3] Steven L. Horowitz, "A syntactic algorithm for peak detection in waveforms with applications to cardiography", May 1975 Communications of the ACM, Volume 18 Issue 5

[4] Hagen Soltau, Tanja Schultz, Martin Westphal, and Alex Waibel. "Recognition of music types", In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-1998). Seattle, Washington, May 1998.

[5] Petri Toiviainen and Tuomas Eerola, "Method for comparative analysis of folk music based on musical feature extraction and neural networks", In III International Conference on Cognitive Musicology, pages 41–45, Jyvskyl, Finland, 2001.

[6] Ponce de Leon, P.J., I˜nesta, J.M. "Feature-driven recognition of music styles. In: 1st Iberian Conference on Pattern Recognition and Image Analysis". LNCS, 2652.(2003) 773–781

[7] McKay, C., Fujinaga, I., "Automatic genre classification using large high-level musi-cal feature sets. In: Int. C. Music Information Retrieval", ISMIR'04. (2004) 525–530

[8] Erling Wold,Thom Blum,Douglass Keislar,James Wheaton , "Content-Based Classification, Search, and Retrieval of Audio", IEEE MultiMedia archive Volume 3 , Issue 3 (September 1996)

[9] 13. Dempster, A.P., Laird, N.M., Rubin, D.B., "Maximum likelihood from incomplete data via the EM algorithm", J. of the Royal Statistical Society B 39 (1977) 1–3

[10] H. Farhat, J. Blacking, "The Dastgah Concept in Persian Music (Cambridge Studies in Ethnomusicology)", by.Cambridge University Press, New edition (July 8, 2004)

[11] N. Darabi, "Music from the viewpoint of digital signals, Implementation of the software Navazande" (English title), Proceeding of 7th Iranian students conference on electrical engineering, 2004.

[12] scala project avialable at http://www.xs4all.nl/~huygensf/scala