

Classification of the Radif of Mirza Abdollah a Canonic Repertoire of Persian Music Using SVM Method

Mahmood Abbasi LAYEGH¹, Siamak HAGHIPOUR^{2,✦}, Yazdan Najafi SAREM³

¹Urmia University Department of Electrical Engineering, Iran

²Tabriz Islamic Azad University, Department of Biomedical Engineering, Iran

³Urmia University of Technology, Faculty of Electrical Engineering, Iran

Received: 20.01.2013 Accepted:11.10.2013

ABSTRACT

Automatic music classification is very useful to music indexing, content-based retrieval and on-line music distribution, but it is a challenge to extract the most common and salient themes from unstructured raw music data. In this paper, a novel approach is proposed to automatically classify the Radif of Mirza Abdollah a canonic repertoire of Persian music. The Radif is made up essentially of non-measured pieces or free rhythm which provide a generative model or pattern for the creation of new composition. Music Segments are decomposed according to time segments obtained from the beginning parts of the original music signal into segments of 3 sec. In order to better classify pure and vocal music, a number of features including inharmonicity, mel-frequency cepstral coefficient, pitch, mean and standard deviation of spectral centroid are extracted to characterize the music content which are mainly related to frequency domain. Experimental results are carried out on a novel database, which contains 250 gushe of the repertoire played by the four most famous Iranian masters and performed on two stringed instruments the Tar & The Setar. Classical machine learning algorithms such as MLP neural networks, KNN and SVM are employed. Finally, SVM shows a better performance in music classification than the others.

Keywords: Repertoire, inharmonicity, Mel-Frequency Cepstral Coefficient, pitch, gushe, K- nearest neighbors, support Vector Machines

1. INTRODUCTION

The creation of huge databases coming from both the restoration of existing analog archives and new content is demanding more and more reliable and fast tools for content analysis and description to be used for searches, content queries, and interactive access. In that context, music genres are crucial descriptors since they have been widely used for years to organize music catalogues, libraries, and music stores. Despite their use, music genres remain a poorly defined concept, which makes the automatic classification problem a nontrivial task. In this article, the state of the art in automatic genre classification is reviewed and new directions in automatic organization of music collections are presented. Saunders [1] proposed a technique for discrimination of audio as speech or

music using the energy contour and the zero-crossing rate. This technique was applied to broadcast radio divided into segments of 2.4 sec which were classified using features extracting from intervals of 16 ms. Four measures of the skewness of the distribution of the zero-crossing rate were used with a 90% correct classification rate by using multivariate Gaussian classifier. When a probability measure on signal energy was added, a performance of 98% is reported. Saad et al [2] have presented a technique to automatically classify audio signal into either speech or music. From 200 frames each of length around 20 ms with adjacent frames overlap one-half of a frame (i.e. 10 ms) they extracted five features: Percentage of Low Energy Frames, Roll Off Point of the Spectrum, Spectral Flux, Zero-Crossing Rate and Spectral Centroid. Scheirer and Slaney used thirteen features in time, frequency and

✦Corresponding author, e-mail: haghypour1@yahoo.com

cepstrum domains and different classification method to achieve a robust performance. Tao Li and Mitsunori, Ogihara used Daubechies Wavelet Coefficient Histograms (DWCH) for music feature extraction for music information retrieval. Carlos N.Silla Jr.et al proposed a non-conventional approach for the automatic music genre classification problem [3]. Changsheng Xu et al proposed effective algorithms to classify and summarized music content [4]. Aggelos Pikariks et al used variable duration hidden Markov models to classify musical patterns [5]. Section 2 introduces the traditional repertoire of Persian music. Section 3 presents preprocessing and proposed method for classification of the radif of Iranian traditional music, and section 4 describes feature extraction and music classification algorithms are used. Section 5 and 6 present conclusions and discussions of future work priorities.

2. PERSIAN MUSIC

2.1. The organization of radif

The radif is the principal emblem and the heart of Persian music, a form of art as quintessentially Persian as that nation's fine carpet and exquisite miniature (Nettl, 1987)[6]. The radif indeed, is not only the musical masterpiece of Iranian genius, but also the artifact which distinguishes its music from all other forms (except the Azerbaijani tradition which has developed its own radif on the same basis). The radif is made up essentially of non-measured pieces («free rhythm») which provide a generative model or pattern for the creation of new compositions, mainly measured, as well as for free improvisation. The radif is a musical

treasure of exceptional richness, the study of which can be approached from different angles such as theory, practice, pedagogy and cultural sociology. In formal terms, it can be defined in the following way: «Radif (which means rank, random, series) actually signifies the order in which the gushe are played; this word also means the totality of the twelve âvâz and dastgâh , as each is played by such and such a master. Thus, the same âvâz can have several radif, each composed or arranged by a different master. The name of radif is best used before the name of a master who arranged it, and sometimes composed junctions between the gushe [6].

3. PROPOSED ALGORITHM FOR PERSIAN RADIF CLASSIFICATION

Only limited related methods have been proposed to automatically classify music. The proposed algorithm is applied on 20 sec. sequences from 250 pieces of music, from different Dastghah and Avaz. The sampling frequency is 44.1KHZ, and the sampling is started after the beginning of the input signal. The 20 sec. music sequences are normalized in order to homogenize the input data for the classifiers. After that the sequences by a factor of 4 are down sampled. Then some features such as inharmonicity, pitch, MFCC, Spectral centroid mean and standard deviation for spectral centroid are calculated for each 3 sec. frame with 90 percent overlap between frames on the input signal. Hamming window is applied to the input signals. Finally support vector machines (SVM) are employed in order to compute the appropriate output. Figure 1 depicts the proposed algorithm.

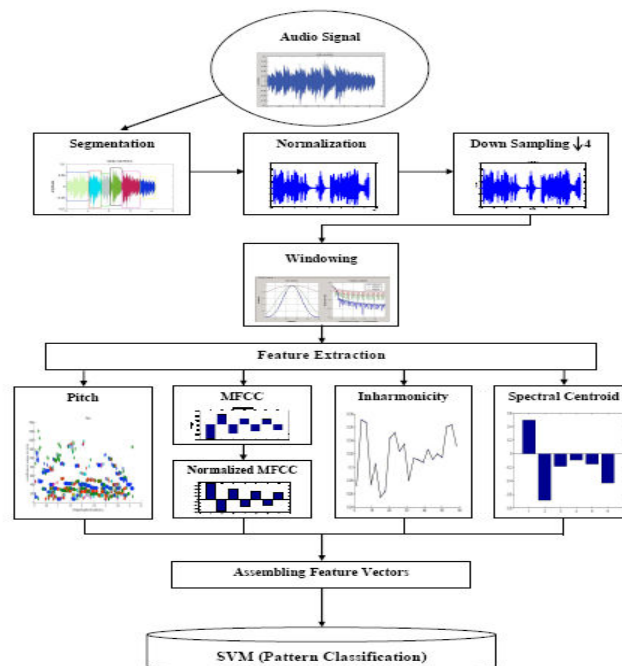


Fig. 1. Proposed algorithm

3.1. Supervised Classifiers

A. Artificial neural networks (ANNs)

ANNs are composed of a large number of highly interconnected processing elements (neurons) jointly working to solve specific problems. The most widely used supervised ANN for pattern recognition is the multilayer perceptron (MLP). It is a very general model that can in principle approximate any nonlinear function. MLPs have been used in the context of artist identification. Neural networks, as well as the other reviewed architectures (except HMMs), can only handle static patterns. This weakness is partly overcome by inputting a number of adjacent feature vectors into the network so that contextual information is taken into account: this strategy corresponds to the so called feed forward time-delay neural network (TDNN) [7].

B. K-nearest neighbor (KNN)

KNN is a nonparametric classifier based on the idea that a small number of neighbors influence the decision on a point. More precisely, for a given feature vector in the target set, the K closest vectors in the training set are selected (according to some distance measures) and the target feature vector is assigned the label of the most represented class in the K neighbor (there is actually no other training than storing the features of the training set) [8-10].

C. Support vector machines (SVMs)

SVMs are based on two properties: margin maximization (which allows for a good generalization of the classifier) and nonlinear transformation of the feature space with kernels (as a data set is more easily separable in a high dimensional feature space). SVMs have been used in the context of genre classification.

C.1 Support Vector Classification

The classification problem can be restricted to consideration of the two-class problem without loss of generality. In this problem the goal is to separate the two classes by a function which is induced from available examples. The goal is to produce a classifier that will work well on unseen examples, i.e. it generalizes well. Consider the example in Figure 2. Here there are many possible linear classifiers that can separate the data, but there is only one that maximizes the margin (maximizes the distance between it and the nearest data point of each class). This linear classifier is termed the optimal separating hyper plane. Intuitively, this boundary is expected to generalize well as opposed to the other possible boundaries [11-14].

C.2 The Optimal Separating Hyper plane

Consider the problem of separating the set of training vectors belonging to two separate classes,

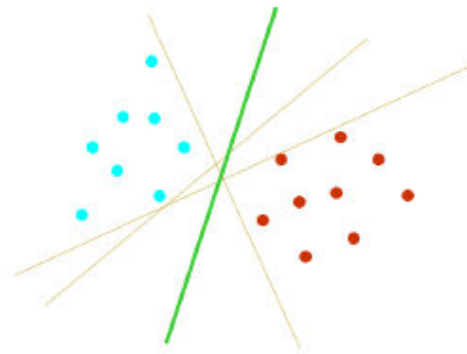


Fig. 2. Optimal Separating Hyperplane

$D = \{(x^1, y^1), \dots, (x^l, y^l)\}, x \in \mathfrak{R}^n, y \in \{-1, 1\}$, with a hyper plane,

$$\langle \omega, x \rangle + b = 0. \tag{1}$$

The set of vectors is said to be optimally separated by the hyper plane if it is separated without error and the distance between the closest vectors to the hyper plane is maximal. There is some redundancy in Equation 1, and without loss of generality it is appropriate to consider a canonical hyper plane (Vapnik, 1995) where the parameters w, b are constrained by,

$$\min_i |\langle \omega, x^i \rangle + b| = 1. \tag{2}$$

This incisive constraint on the parameterization is preferable to alternatives in simplifying the formulation of the problem. In words it states that: the norm of the weight vector should be equal to the inverse of the distance, of the nearest point in the data set to the hyper plane. The idea is illustrated in Figure 3, where the distance from the nearest point to each hyper plane is shown. A separating hyper plane in canonical form must satisfy the following constraints,

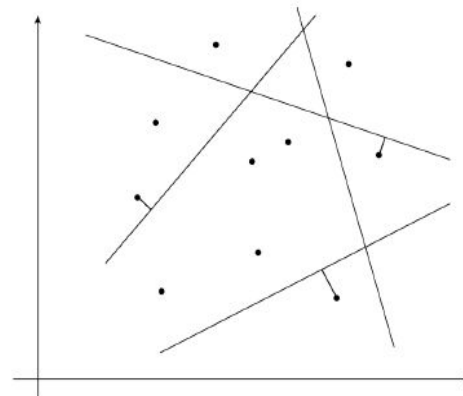


Fig. 3. Canonical Hyperplanes

$$y^i [\langle \omega, x^i \rangle + b] \geq 1, i = 1, \dots, l. \tag{3}$$

The distance $d(w, b; x)$ of a point x from the hyperplane (w, b) is,

$$d(w, b; x) = \frac{|\langle w, x^i \rangle + b|}{\|w\|} \quad (4)$$

The optimal hyper plane is given by maximising the margin, ρ , subject to the constraints of Equation 2 The margin is given by,

$$\begin{aligned} \rho(w, b) &= \min_{x^i, y^i = -1} d(w, b; x^i) + \min_{x^i, y^i = 1} d(w, b; x^i) \\ &= \min_{x^i, y^i = -1} \frac{|\langle w, x^i \rangle + b|}{\|w\|} + \min_{x^i, y^i = 1} \frac{|\langle w, x^i \rangle + b|}{\|w\|} \\ &= \frac{1}{\|w\|} \left(\min_{x^i, y^i = -1} |\langle w, x^i \rangle + b| + \min_{x^i, y^i = 1} |\langle w, x^i \rangle + b| \right) \\ &= \frac{2}{\|w\|} \end{aligned} \quad (5)$$

Hence the hyper plane that optimally separates the data is the one that minimizes

$$\Phi(w) = \frac{1}{2} \|w\|^2 \quad (6)$$

It is independent of b because provided Equation 3 is satisfied (i.e. it is a separating hyper plane) changing b will move it in the normal direction to itself. Accordingly, the margin remains unchanged but the hyper plane is no longer optimal in that it will be nearer to one class than the other. To consider how minimizing Equation 6 is equivalent to implementing the SRM principle, suppose that the following bound holds,

$$\|w\| < A. \quad (7)$$

Then from Equation 3 and 4,

$$d(w, b; x) \geq \frac{1}{A} \quad (8)$$

Accordingly the hyper planes cannot be nearer than $\frac{1}{A}$ to any of the data points and intuitively it can be seen in Figure 4 how this reduces the possible hyper planes, and hence the capacity. The VC dimension, h , of the set of canonical hyper planes in n dimensional space is bounded by,

$$h \leq \min[R^2 A^2, n] + 1 \quad (9)$$

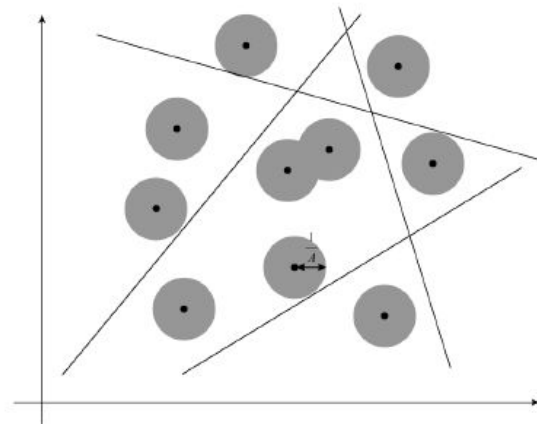


Fig. 4. Constraining the Canonical Hyper planes

where R is the radius of a hypersphere enclosing all the data points. Hence minimizing Equation 6 is equivalent to minimizing an upper bound on the VC dimension. The solution to the optimization problem of Equation 5 under the constraints of Equation 3 is given by the saddle point of the Lagrange functional (Lagrangian) (Minoux, 1986),

$$\Phi(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y^i [\langle w, x^i \rangle + b] - 1) \quad (10)$$

where α are the Lagrange multipliers. The Lagrangian has to be minimized with respect to w, b and maximized with respect to $\alpha \geq 0$. Classical Lagrangian duality enables the primal problem, Equation 10, to be transformed to its dual problem, which is easier to solve. The dual problem is given by,

$$\max_{\alpha} W(\alpha) = \max_{\alpha} \left(\min_{w, b} \Phi(w, b, \alpha) \right) \quad (11)$$

The minimum with respect to w and b of the Lagrangian, Φ , is given by,

$$\begin{aligned} \frac{\partial \Phi}{\partial b} = 0 &\Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \\ \frac{\partial \Phi}{\partial w} = 0 &\Rightarrow w = \sum_{i=1}^l \alpha_i y_i x_i \end{aligned} \quad (12)$$

Hence from Equations 10, 11 and 12, the dual problem is,

$$\max_{\alpha} W(\alpha) = \max_{\alpha} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{k=1}^l \alpha_k \quad (13)$$

and hence the solution to the problem is given by,

$$\alpha^* = \arg \min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{k=1}^l \alpha_k \quad (14)$$

with constraints,

$$\alpha_i \geq 0 \quad i=1, \dots, l$$

$$\sum_{j=1}^l \alpha_j y_j = 0 \quad (15)$$

Solving Equation 14 with constraints Equation 15 determines the Lagrange multipliers, and the optimal separating hyperplane is given by,

$$\omega^* = \sum_{i=1}^l \alpha_i y_i x_i \quad (16)$$

$$b^* = -\frac{1}{2} \langle \omega^*, x_r + x_s \rangle$$

where x_r and x_s are any support vector from each class satisfying,

$$\alpha_r, \alpha_s > 0, \quad y_r = -1, \quad y_s = 1 \quad (17)$$

The hard classifier is then,

$$f(s) = \text{sgn}(\langle \omega^*, x \rangle + b) \quad (18)$$

Alternatively, a soft classifier may be used which linearly interpolates the margin,

$$f(s) = h(\langle \omega^*, x \rangle + b) \quad \text{where } h(z) = \begin{cases} -1 & : z < -1 \\ z & : -1 \leq z \leq +1 \\ +1 & : z > +1 \end{cases} \quad (19)$$

This may be more appropriate than the hard classifier of Equation 18, because it produces a real valued output between -1 and 1 when the classifier is queried within the margin, where no training data resides. From the Kuhn-Tucker conditions,

$$\alpha_i (y^i [\langle \omega, x^i \rangle + b] - 1) = 0, \quad i = 1, \dots, l, \quad (20)$$

and hence only the points x^i which satisfy,

$$y^i [\langle \omega, x^i \rangle + b] = 1 \quad (21)$$

will have non-zero Lagrange multipliers. These points are termed Support Vectors (SV). If the data is linearly separable all the SV will lie on the margin and hence the number of SV can be very small. Consequently the hyper plane is determined by a small subset of the training set; the other points could be removed from the training set and recalculating the hyper plane would produce the same answer. Hence SVM can be used to summarize the information contained in a data set by the SV produced. If the data is linearly separable the following equality will hold,

$$\|\omega\|^2 = \sum_{i=1}^l \alpha_i = \sum_{i \in SV_s} \alpha_i = \sum_{i \in SV_s} \sum_{j \in SV_s} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle. \quad (22)$$

Hence from Equation 11 the VC dimension of the classifier is bounded by,

$$h \leq \min[R^2 \sum_{i \in SV_s}^l, n] + 1, \quad (23)$$

and if the training data, x , is normalized to lie in the unit hypersphere,

$$h \leq 1 + \min \sum_{i \in SV_s}^l, n], \quad (24)$$

4. FEATURE EXTRACTION

The first step in a classification problem is typically data reduction. The data reduction stage, which is also called feature extraction, consists of a few important facts about each class. Since audio data contains much redundancy, important features are lost in the dissonance of unreduced data. The choice of features is critical as it greatly affects the accuracy of audio classification. The selected features must reflect the significant characteristics of each class of audio signals.

Features related to spectral domain are considered. Typically, audio features are extracted in two levels: short-term frame-level and long-term clip-level. Here, a frame is defined as a group of adjacent samples lasting from 10ms to 40ms. The audio signal within such periods presumably remains stationary and short-term features both in time-domain and frequency domain can be extracted. The features studied in this research include: inharmonicity, Mel-frequency cepstral coefficient, pitch, mean of spectral centroid and standard deviation of centroid[15].

4.1. Inharmonicity

In music, inharmonicity is the degree to which the frequencies of overtones (known as partials, partial tones, or harmonics) depart from whole multiples of the fundamental frequency. Acoustically, a note perceived to have a single distinct pitch in fact contains a variety of additional overtones. Many percussion instruments, such as cymbals, tam-tams, and chimes, create complex and inharmonic sounds. However, in stringed instruments such as the piano, violin, guitar, t r and set r, the overtones are close to or in some cases, quite exactly whole number multiples of the fundamental frequency. Any departure from this ideal harmonic series is known as inharmonicity. The less elastic the strings are (that is, the shorter, thicker, and stiffer they are), the more inharmonicity they exhibit. Figure 5 shows inharmonicity coefficients for Dastgah-e Mahur(Gushe : Araq).[16, 17].

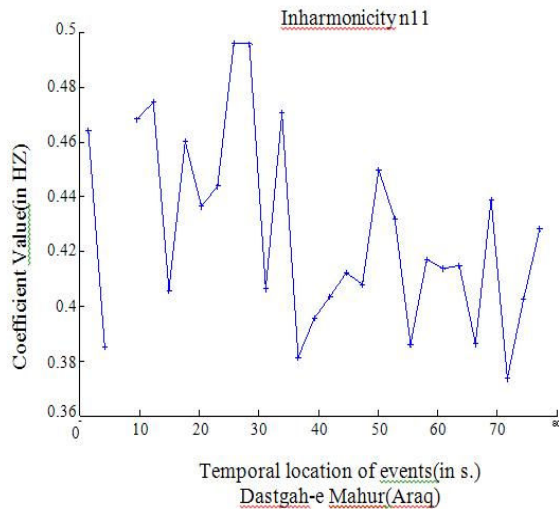


Fig. 5. Inharmonicity for for Dastgah-e Mahur(Gushe : Araq)

4.2. Pitch

In music, the position of a tone in the musical scale is designated by a letter name and determined by the frequency of vibration of the source of the tone. Pitch is an attribute of every musical tone. The fundamental or first harmonic of any tone is perceived as its pitch. Absolute pitch is the position of a tone in the musical scale determined according to its number of vibrations per second, irrespective of other tones. The term also denotes the capacity to identify any tone upon hearing it sounded alone or to sing any specified tone. For example pitch helps the human ear to distinguish between string instruments, wind instruments and percussion instruments such as the drums, tabla etc. Figure 6 shows Pitch coefficients for Dastgah-e Nava(Gushe : Naghme) [18,19].

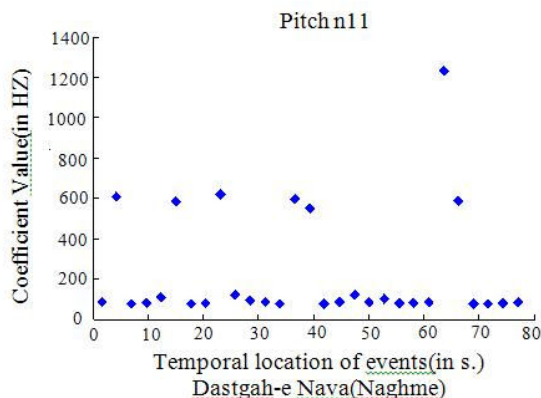


Fig. 6. Pitch for Dastgah-e Nava(Gushe : Naghme)

4.3. Mel-Frequency Cepstral Coefficients

MFCCs are a short-time spectral decomposition of an audio signal that conveys the general frequency

characteristics important to human hearing. MFCCs are commonly used in the field of speech recognition. Recent research shows that MFCCs are capable of capturing useful sound characteristics of music files as well. The premise is that MFCC's contain enough information about the timbre of a song to perform genre classification. To compute these features, a sound file is subdivided into small frames of about 20 ms each and then MFCCs are computed for each of these frames. Since the MFCCs are computed over short intervals of a song, they do not carry much information about the temporal attributes of a song, such as rhythm or tempo.

Process for converting waveforms to MFCCs

The general process for converting a waveform to its MFCCs is described in Logan and roughly explained by the following pseudo code:

1. Take the Fourier Transform of a frame of the waveform.
2. Smooth the frequencies and map the spectrum obtained above onto the mel scale.
3. Take the logs of the powers at each of the mel frequencies.
4. Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the resulting spectrum[20, 21].

N MFCC coefficients are computed for all short duration frames of a wav file and store them in a FX_N size matrix, where F is the number of frames.

The main purpose of the MFCC is to imitate the behavior of a human ear. Psychophysical studies have shown that human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency f , measured in Hz, a subjective pitch is measured on a scale called the Mel scale. The Mel-frequency is linear frequency spacing below 1 kHz and a logarithmic spacing above 1 kHz.

Therefore, filters spaced linearly at low frequency and logarithmic at high frequencies can be used to capture the phonetically important characteristics (voiced and unvoiced) of the speech. The commonly used formula to approximately reflect the relation between the Mel-frequency and the physical frequency is given by:

$$DCT(\log_{10}(\text{abs}(FFT(\text{signal})))) \Rightarrow MFCCs \quad (25)$$

Mel-frequency cepstral coefficients (MFCC) are perceptually motivated features that are also based on the STFT (Short Time Fourier Transform), after taking the log-amplitude of the magnitude spectrum. The FFT (Fast Fourier Transform) bins are grouped and smoothed according to the perceptually motivated Mel-frequency scaling. Finally, in order to decorrelate the resulting feature vectors. A Discrete Cosine Transform (DCT) is performed. Although typically 12 coefficients are used for speech representation, have found that the first five coefficients provide the best classification performance, the first seven bins, and compute the

mean and variance of each over the frames are used. Figure 7 shows different steps of calculating MFCC.

Seven coefficients for Mel frequency for Dastgah-e Mahur(Gushe : Dad) are shown in Figure 8.

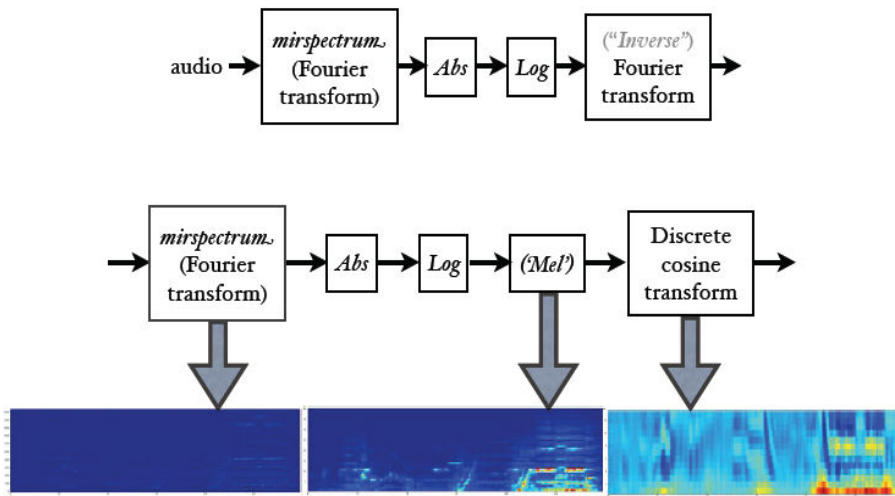


Fig.7. The scheme for calculating MFCC

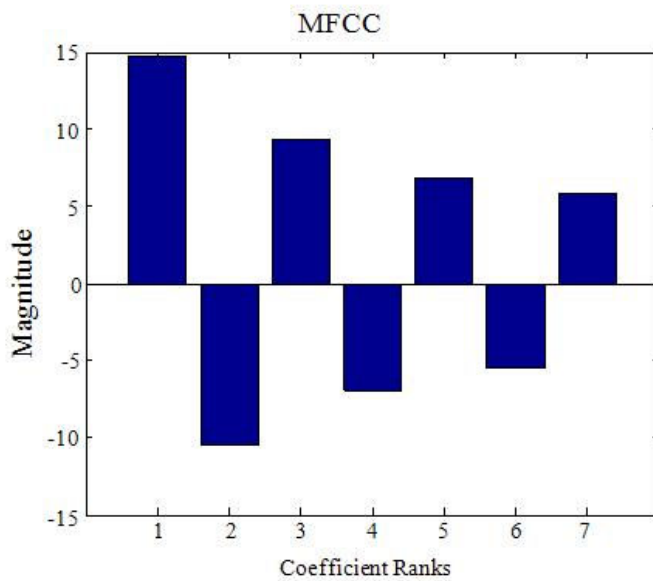


Fig. 8. MFCC coefficients For Dastgah-e Mahur(Gushe : Dad)

4.4. Mean and standard deviation of spectral centroid

The spectral centroid of a sound is a concept adapted from psychoacoustics and music cognition. It measures the average frequency, weighted by amplitude, of a spectrum. In cognition applications, it is usually averaged over time. The standard formula for the (average) spectral centroid of a sound is:

$$C = \frac{\sum_j c_j}{j} \tag{26}$$

Where c_j is the centroid for one spectral frame, and i is the number of frames for the sound. A spectral frame is some number of samples which is equal to the size of the FFT. The (individual) centroid of a spectral frame is defined as the average frequency weighted by amplitudes, divided by the sum of the amplitudes or:

$$c_j = \frac{\sum_j f_j a_j}{\sum_j a_j} \tag{27}$$

In practice, *Centroid* finds this frequency for a given frame, and then finds the nearest spectral bin for that frequency. This yields an accuracy which is dependent upon the analysis frame size. The centroid is usually a

lot higher than one might intuitively expect, because there is so much more energy above (than below) the fundamental which contributes to the average. Care should be taken not to confuse the centroid with the fundamental. Rather, it is often used as a measure of "brightness" in comparing sounds. The higher the centroid, the "brighter" the sound. Time-varying centroids like those in our implementation, are uncommon in the literature, but can be useful in comparing sound morphologies, or in this case, in designing a kind of experimental filter [22]. MFCC, Mean and standard deviation of spectral centroid for Dastgah-e Rast-Panjgah(Gushe:Neyriz) are shown in Figure 9.

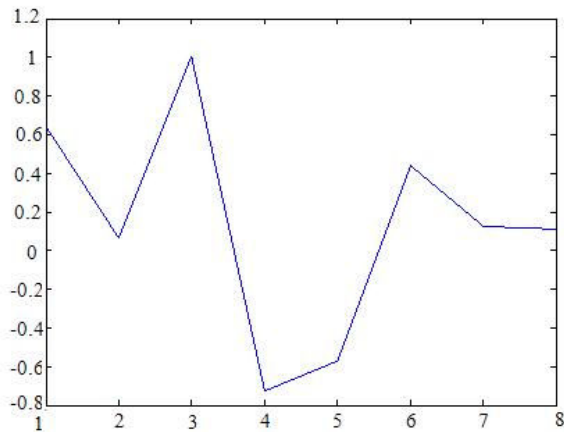


Fig. 9. Mean and standard deviation of spectral centroid for Dastgah-e Rast-Panjgah(Gushe : Neyriz)

5. EXPERIMENTAL RESULTS

The music dataset used in musical genre classification of this experiment contains 1250 music samples. It consists of long music files covering the 12 modal systems (Dastghah and Avaz) of Persian music. The files were collected from 5 CDs recorded with the collaboration of several well-known masters of the târ and the setâr. For classification of the Radif of Mirza Abdollah, a number of features such as inharmonicity, mel-frequency cepstral coefficient (MFCC), pitch, mean and standard deviation of spectral centroid are utilized. Then a number of commonly used supervised machine learning algorithms for music genre classification such as multilayer perceptron (MLP), K-nearest neighbor (KNN) and Support vector machines (SVMs) are used. Table 1 compares the classification accuracy extraction MLP and KNN. Table 2 compares the classification accuracy using SVM with four different Kernel functions. Experiments show support vector machine learning method has better performance in musical genre classification and is more advantageous than traditional Euclidean distance based method and other statistic learning methods. As can be seen from Table 2, RBF support vector machine classifier far outweighs other kernel functions in the performance and Figure 10 shows the percentage of the results of various Kernel functions for the Radif of Mirza Abdollah in different âvâz and dastgâh.

Table 1. Feature extraction using MLP & KNN

Dastgah & Avaz	DATA Num	MLP	KNN
Dastgah-e mahur	44	0.2000	0.2170
Dastgah-e shour	34	0.2000	0.1000
Dastgah-e Segah	20	0.3125	0.0625
Dastgah-e Chahargah	31	0.2000	0.3333
Dastgah-e Rast Panjgah	25	0.0900	0.1370
Dastgah-e homayoon	28	0.1900	0.0480
Dastgah-e Nava	21	0.3300	0.2500
Avaz-e Bayat-e Turk	15	0.0900	0.0900
Avaz-e AabuAta	10	0.3333	0.3333
Avaz-e Afshari	4	0.6667	0.6667
Avaz-e Bayat-e Esfahan	7	0.3333	0.1667
Avaz-e dashti	5	0.6667	0.6670
Avaz -e Kord-e Bayat	8	0.8000	0.2000

Table 2. Feature extraction using SVM

Dastgah & Avaz	Kernel Function			
	Linear	Quadratic	Poly Order 3	RBF sigma=1
Dastgah-e mahur	0.9309	0.9463	0.9655	0.9706
Dastgah-e shour	0.9310	0.9483	0.9379	0.9655
Dastgah-e Segah	0.9348	0.9484	0.9076	0.9565
Dastgah-e Chahargah	0.9152	0.9182	0.9273	0.9545
Dastgah-e Rast Panjgah	0.9276	0.9290	0.9091	0.9688
Dastgah-e homayoon	0.9309	0.9094	0.9048	0.9677
Dastgah-e Nava	0.9320	0.9250	0.9056	0.9620
Avaz-e Bayat-e Turk	0.8693	0.8693	0.7955	0.9375
Avaz-e AabuAta	0.6250	0.7292	0.7500	0.8750
Avaz-e Afshari	0.4444	0.5556	0.4444	0.6667
Avaz-e Bayat-e Esfahan	0.6875	0.7500	0.7292	0.8750
Avaz-e dashti	0.4167	0.6667	0.4167	0.8333
Avaz –e Kord-e Bayat	0.5714	0.7429	0.6000	0.8571

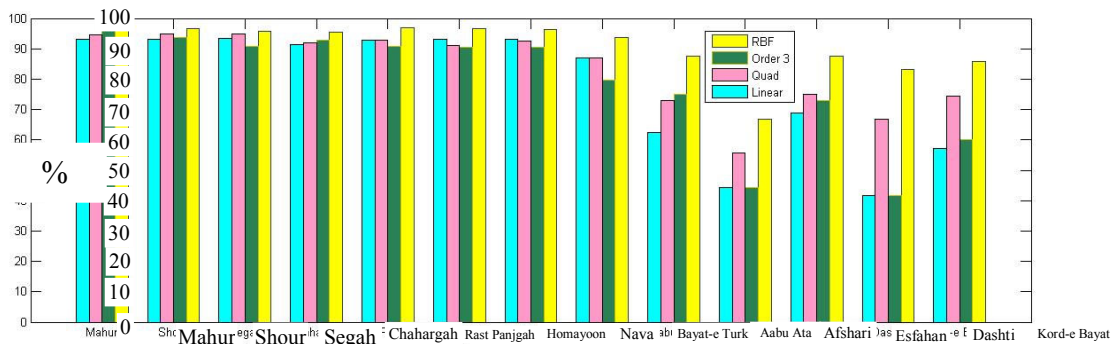


Fig. 10 SVM various kernel function results

6. CONCLUSIONS AND FUTURE WORK

Many techniques have been proposed in the literature for music classification .In order to achieve acceptable performance,most of them require a large amount of training data.In this paper extensive experimentation on a diverse set of audio data using different classification frameworks and introducing new features for Iranian traditional music has been conducted. The results for 3sec duration audio samples clearly show that RBF networks give satisfactory results for the radif of Mirza Abdollah since RBF networks depend on the centers of clusters.The proposed algorithm has shown inspiring improvements in the results that that indicate the viability of the approach. There are two directions that need to be investigated in the future. The first direction is to improve the computational efficiency for support vector machines. Support vector machines take a long time in the training process, especially with a large

number of training samples. Therefore, how to select proper kernel function and determine the relevant parameters is extremely important. The second direction is to make the classification result more accurate. To achieve this goal, it is mandatory to explore more music features that can be used to characterize the music content.

REFERENCES

[1] Seibt P., “Algorithmic Information Theory”, Mathematics of Digital Information Processing, pp. 23-30. *Springer*, New York (2006)

[2] Meana H.B., “Advances in Audio and Speech Signal Processing” *Technologies and Applications Idea Group*, Mexico (2007)

- [3] Li T., Ogihara M. "Toward Intelligent Music Information Retrieval", *IEEE Trans. Multimedia*, vol. 13, no. 3. pp. 564-574 (2005)
- [4] Changsheng X., "Automatic Music Classification and Summarization", *IEEE transactions on speech and audio processing*. Vol. 13. No. 3. pp. 441-450 (2005)
- [5] Pikarkis A., Theoridis S., Kamarotos D. "Classification of Musical Patterns Using Variables Duration Hidden Markov Models" *IEEE Transactions on Audio Speech and Language Processing*, Vol.14. No.5. p.p. 1795-1807 (2006)
- [6] Simms R., Koushkani A., "Mohammad Reza Shajarian's avaz in Iran and beyond 1979 – 2010" *Lexington Books*, United States of America (2012)
- [7] Scaringella N., Zoia G., Mlynek D., "Automatic Genre Classification of Music Content" *IEEE signal processing*, 23(2): 133-141 (2006)
- [8] Yu Q., Michel Y., Sorjamaal A., Guillen A., Lendasse A., Severin E., "OP-KNN Method and Applications" Hindawi (2010)
- [9] Zhu J., Xue X., Lu L., "Musical Genre Classification By Instrumental Features", Paper presented on the International Computer Music Conference, Shanghai, China, ICMC 2004
- [10] Cunningham P., Delany S., "k-Nearest Neighbour Classifiers", *Technical Report UCD-CSI-2007*. March 27,(2007)
- [11] Steve R.G., "Support Vector Machines for Classification and Regression", *University Of Southampton*, Technical Report, 10 May (1998)
- [12] Carlos N., Silla J., Alessandro L., Kaestner A., "A Machine Learning Approach to Automatic Music Genre Classification", *Brazilian Computer Society*, Brazil (2008)
- [13] Hao Zhang Alexander C. Berg Michael Maire Jitendra Malik, SVM-KNN, "Discriminative Nearest Neighbor", Paper presented on the Computer Vision and Pattern Recognition, Rhode Island, 17-22 June (2006)
- [14] Changsheng X., Maddage N.C., Shao X., Cao F., Tian Q., "Musical Genre Classification Using Support Vector Machines", Paper presented on the 2003 IEEE International Conference. Singapore, 6-10 April 2003
- [15] Saeed Khan M.K., "Automatic Classification of Speech & Music in Digitized Audio", Master degree thesis, *University of Dhahran* (2005)
- [16] Galembo A., Askenfelt A., "Measuring inharmonicity through pitch extraction", Quarterly Progress and Status Report, STL-QPSR. vol. 35, pp. 135-144 (1994)
- [17] Järveläinen H., Välimäki V., Karjalainen M., *Laboratory of Acoustics and Audio Signal Processing*, Doi: 10.1121/1.1374756
- [18] Zhou R., "Feature Extraction of Musical Content for Automatic Music Transcription", Master Thesis, *Chinese Academy of Science* (2006)
- [19] Herman G.L., "Fundamental Frequency Tracking In Music With Multiple Voices", Thesis for the degree of Master of Science., *University of Illinois at Urbana-Champaign* (2007)
- [20] Jiajun Z., Xiangyang X., Hong L., "Musical Genre Classification By Instrumental Features", paper presented on International Computer Music Conference, *Fudan University*, Shanghai. China (2004)
- [21] Aliaksandr V., Paradzinets., "Variable Resolution Transform-based Music Feature Extraction and their Applications for Music Information Retrieval", Dissertation, *University of Lyon* (2007)
- [22] Tzanetakis G., Essl G., Cook P., "Automatic Musical Genre Classification Of Audio Signals", *IEEE Transactions On Speech And Audio Processing*, Vol. 10. No. 5. 293-302 (2002).