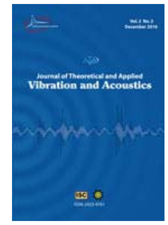
Journal of Theoretical and Applied Vibration and Acoustics 2(2) 107-118 (2016)

I S A V

Journal of Theoretical and Applied

Vibration and Acoustics

journal homepage: http://tava.isav.ir



Classification of Iranian traditional musical modes (DASTGÄH) with artificial neural network

Borhan Beigzadeh *, **Mojtaba Belali Koochesfahani**

Biomechatronics and Cognitive Engineering Research Lab, School of Mechanical Engineering, Iran University of Science and Technology, Tehran, Iran

A R T I C L E I N F O

Article history:
Received 17 December 2015

Received in revised form
9 May 2016

Accepted 15 May 2015

Available online 06 July 2016

Keywords:
Iranian traditional musical modes (DASTGÄH)

Classification

Artificial Neural Network

Feature extraction

A B S T R A C T

The concept of Iranian traditional musical modes, namely DASTGÄH, is the basis for the traditional music system. The concept introduces seven DASTGÄHs. It is not an easy process to distinguish these modes and such practice is commonly performed by an experienced person in this field. Apparently, applying artificial intelligence to do such classification requires a combination of the basic information in the field of traditional music with mathematical concepts and knowledge. In this paper, it has been shown that it is possible to classify the Iranian traditional musical modes (DASTGÄH) with acceptable errors. The seven Iranian musical modes including SHÖR, HOMÄYÖN, SEGÄH, CHEHÄRGÄH, MÄHÖR, NAVÄ and RÄST-PANJGÄH are studied for the two musical instruments NEY and Violin as well as for a vocal song. For the purpose of classification, a multilayer perceptron neural network with supervised learning method is used. Inputs to the neural network include the top twenty peaks from the frequency spectrum of each musical piece belonging to the three aforementioned categories. The results indicate that the trained neural networks could distinguish the DASTGÄH of test tracks with accuracy around 65% for NEY, 72% for violin and 56% for vocal song.

1. Introduction

Music and speech recognition as well as distinguishing different tones or speeds of music is usually possible for an experienced person; however, it requires more skill in the field of music in order to recognize music styles. Essentially, determining whether a piece of (Iranian)

*Corresponding author.
E-mail address: b_beigzadeh@iust.ac.ir (B. Beigzadeh)

http://dx.doi.org/ 10.22064/tava.2016.20515

traditional music is played in a certain Iranian traditional musical mode (DASTGÄH) is performed by human. Few research works have used artificial intelligence for this purpose. However, some researchers have tried to recognize speech from music [1, 2] and some others have performed research to classify pieces of music according to different music styles [3-5].

In some papers, researchers have investigated the use of pattern recognition techniques for recognizing two features of Persian music, DASTGÄH and MAQÄM using their relation to the musical scale and mode [6]. For this purpose, they have defined some statistical measures that characterize the melodic pattern of a musical signal. Finally, they have reported the results for HOMÄYÖN with an average error rate of 28.2%. In [7], fuzzy logic type 2 as the basic part of the system has been used for modeling the uncertainty of tuning the scale steps of each DASTGÄH. Although they could obtain the overall accuracy of 85%, the number of DASTGÄHs were limited to five. In [8], they have used SVM to classify the DASTGÄHs played with TÄR and SETÄR; the results showed accuracy of 45-90% for different DASTGÄHs. An artificial neural network has been used in [9] to classify Persian Musical DASTGÄHs which uses 24 input units (each for one of the possible notes in an octave of Persian music) and 5 hidden units with Gaussian activation functions. This network is able to classify 83.33% of the presented scales correctly. Some researchers have tried to distinguish the scale of MÄHÖR from those of other six DASTGÄHs in the SETÄR instrument using artificial neural network [10]. Furthermore, in other works, numerical patterns are assigned to each of the Iranian musical scales. Thus for a given piece of music, the scale is determined by comparison with these patterns [11]. The proposed process is performed by human, not in an automated manner.

The main purpose of the current study is to distinguish and categorize the DASTGÄH of each played music by NEY and Violin or any vocal music and classify it as one of the seven DASTGÄHs of SHÖR, HOMÄYÖN, SEGÄH, CHEHÄRGÄH, MÄHÖR, NAVÄ and RÄST-PANJGÄH. For this purpose, the artificial neural network is used and in order to train and test the network, the frequency spectrum of the played pieces of music are utilized. As we will conclude in this study, the information of each piece of music regarding its corresponding DASTGÄH, is partially embedded in its frequency spectrum. However, the whole information of the DASTGÄH is not absolutely related to the frequency content of the played music and the observed error corresponding to this method is directly related to this fact.

In the following section, the theory of music is shortly discussed. Then the features used in the classification of musical styles are introduced. Afterward, the approach of the paper in utilizing the Fast Fourier Transform and artificial neural network to classify the musical pieces will be discussed. Finally, the results will be expressed.

2. Music theory

In this section, we introduce and explain the concepts and terms related to the music theory [12] and used in this study.

Frequency: The number of wave cycles occurring in one second is measured as the frequency of the wave. Sounds heard by human range from 20 Hz to 20000 Hz.

Music: Vocal or instrumental sounds combined to produce harmony and beauty of form.

Pitch: An auditory sensation through which a listener assigns musical tones to relative positions on a musical scale primarily based on their perception of the vibration frequency [13].

Tone: A musical tone is a steady periodic sound that is characterized by its duration, pitch, loudness and quality [14]. In this article, the term "tone" refers to the musical interval named "PARDEH" in Persian music.

Note: In music, a note usually refers to a unit of fixed pitch that has been given a name. A note is a discretization of a musical or sound phenomenon and thus facilitates musical analysis [15].

GÖSHEH: Some melodic sections of various lengths are called GÖSHEH. The GÖSHEH can be from less than a half minute to several minutes long. They must follow a specific tone interval and have a SHÄHED (witness note) that is the center around which the melody evolves, the note to which melodic passages constantly return.

DASTGÄH: It is a musical modal system in traditional Persian art music which includes GÖSHEH of the near pitches.

Seven DASTGÄHs: Persian art music consists of seven principal musical modal systems or DASTGÄHs some of which have sub-branches called ÄVÄZ. Table 1 shows the number of DASTGÄHs and their corresponding ÄVÄZes.

Table 1. Iranian traditional musical modes (DASTGÄH)

DASTGÄH	ÄVÄZ
MÄHÖR	
SHÖR	ABÖ-ATÄ, BAYÄT-TURK, AFSHÄRI, DASHTI
SEGÄH	
CHEHÄRGÄH	
HOMÄYÖN	BAYÄT-ESFAHÄN
NAVÄ	
RÄST-PANJGÄH	

3. Features associated with classical music classification

For intelligent classification of DASTGÄHs with the best performance, it is necessary to extract features that are varied in each DASTGÄH. Features for recognition of music style can be divided into three main categories [1, 3, 4]; the feature of timbral texture, the feature of rhythmic content and the feature of pitch content.

Timbre is an agent that causes differences in the voice of a particular note in the various instruments. It is related to the number and relative strength of notes played by different instruments [10].

The feature of rhythmic content is not efficient in the Iranian traditional musical modes (DASTGÄH) recognition. This is because all of these DASTGÄHs can be played with different rhythms.

The feature of pitch content contains melodic and harmonic information about the music. It is possible to find good classification methods by investigating this feature.

In other words, when the same shape is found repeated in a waveform, a source can be identified and characterized by the repetition rate (pitch) and the vibration pattern (timbre) [16].

In the Iranian traditional music, DASTGÄH recognition is partly based on the pitch content [10]. Therefore, the raw data processing includes finding the spectrum which gives the dominant frequency components of the music pieces.

4. Fast Fourier Transform (FFT)

FFT is the main tool in this study to extract the features of the captured and saved music pieces. The neural network is then trained using the results obtained from FFT analysis.

5. Artificial neural network

A perceptron neural network with two hidden layers is used in this work. Moreover, in this study, a feed-forward and backward perceptron neural network is used with supervised learning (see [17] for more detailed study of the concept).

6. Methodology

Arrangement of notes used in the music track, i.e. frequency intervals of a piece of music with respect to the “base note”, determines the musical DASTGÄH [18]. Such definition is an engineering (or mathematical) representation of the “musical DASTGÄH”. According to this definition, the investigation should be conducted on the frequency content of the musical tracks.

In this paper, the purpose is to design a system to receive an Iranian music track as input, and identify its corresponding DASTGÄH employing a neural network. In this regard, we should first train the neural network by short simple music tracks in various DASTGÄHs. We try to set no kind of restrictions and limitations to the input tracks. However, it was concluded that not every musical track is suitable for DASTGÄH recognition. The limitations are described in the following section.

6.1. Conditions and restrictions

The conditions and restrictions on providing inputs to the neural network model include:

- The two musical instrument “NEY” and “Violin” are played separately in solo mode.
- The vocal traditional song is sung in solo mode.
- Each of the musical tracks is played in just one GÖSHEH of a specific DASTGÄH.
- Each GÖSHEH of a specific DASTGÄH is considered only among that DASTGÄH.
- The vocal music has been sung by just one singer during a specific DASTGÄH.

- Each musical track is recorded for 10 to 20 seconds (minimum time required for DASTGÄH recognition by human [10]).
- Musical tracks which were played by masters of NEY and violin instruments, or were sung by Iranian vocal master used in this project were recorded in the “wave” format. The sampling frequency is 44100 Hz, the rate is 8 bits per second and recording is in the Mono mode.

6.2. Intuitional comparison

FFT analysis was performed to obtain the spectrum of musical tracks. A quick investigation of the spectrums reveals that the musical tracks in a specific DASTGÄH are almost similar in terms of the contents of their spectrums. For example, the frequency spectrum of three musical tracks played by NEY in CHEHÄRGÄH are shown in Fig. 1 for visual comparison.

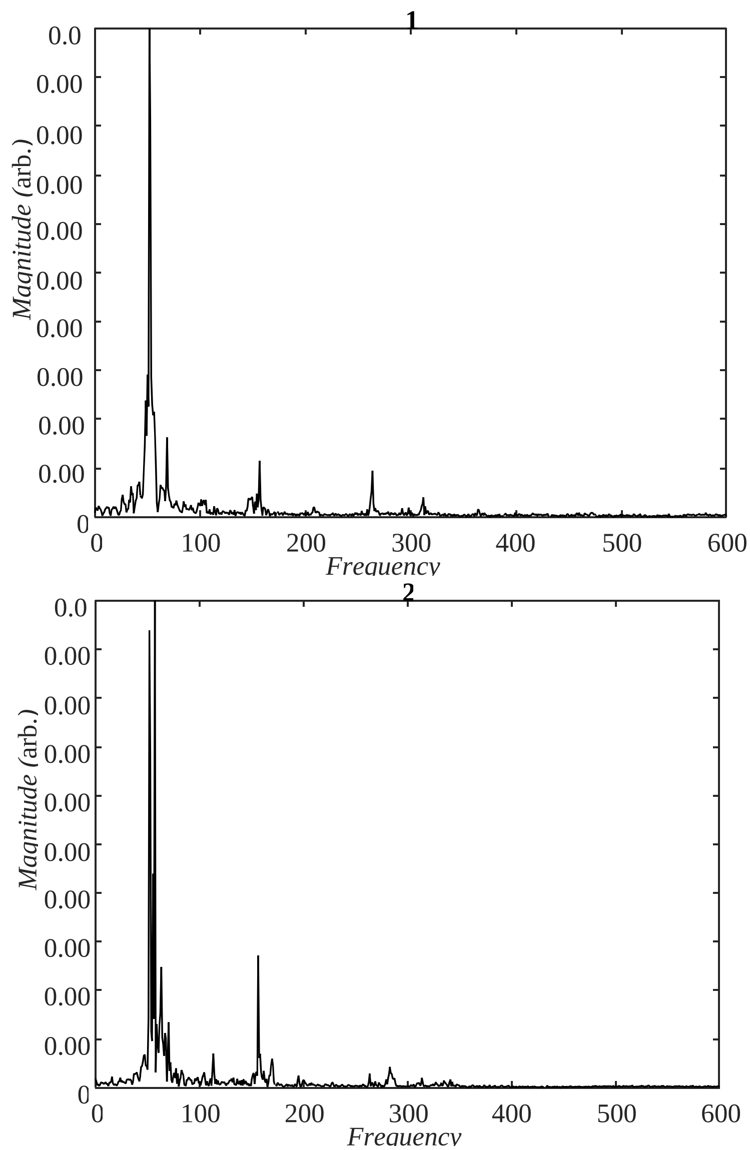


Fig. 1: Spectrum of 3 sound files in CHAHÄRGÄH played with NEY

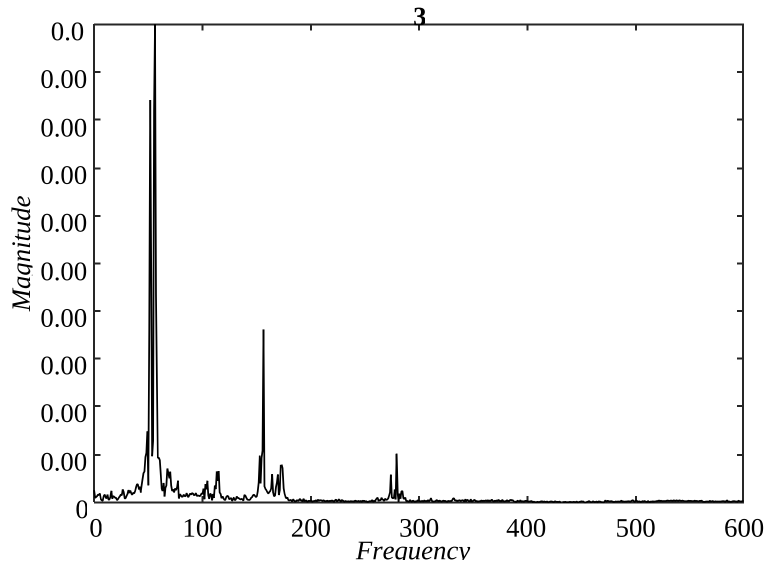


Fig. 1. (Continued)

It was observed that the NEY and violin instruments are more comparable considering their frequency spectrums. However, the accuracy of the spectrum associated with the vocal sound track is not enough comparable with the other two. The reason is that the recorded sound tracks are from an old famous Iranian singer and therefore, the recorded voices embed environmental acoustic noise. Moreover, the degree of inharmonicity (the degree to which the frequencies of overtones depart from integer multiples of the fundamental frequency; a note perceived to have a single distinct pitch in fact contains a variety of additional overtones) in a vocal song is much more than a musical track played with a stringed instrument [8]. It is then expected that the classification of the vocal sound tracks would not be as successful as those associated with NEY and violin; results will support this fact as well.

6.3. Frequency spectrum extraction

FFT analysis was performed on the audio files and all the local maxima associated with 5001 frequency intervals were extracted from the obtained values. This number, i.e. 5001, is not suitable to be the number of features which would be considered as the inputs of the neural network. Therefore, the whole frequency range was divided into 20 windows and the absolute maximum value for each window was extracted. For each track, these 20 values are considered as features to be trained to the ANN. The data was stored in two matrices and finally used as features in the neural network; one matrix as the testing data and the other for the training data.

6.4. Supervised classification of testing and training data

In a supervised learning manner, the data should be divided into two sets of training and testing data. In this direction, a random process is taken into account to determine the testing data to be 30% and training data to be 70% of the data pool. The DASTGÄH for the testing and training data are known since the learning process is a supervised one; manual classification is performed by (an expert) user.

6.5. Neural network learning and error

After normalization of data, network's structure is built and trained by the training data. The testing process is performed by examination of the testing data. The process is to determine the DASTGÄH associated with each testing data.

To calculate the amount of neural network error, musical tracks classified by the neural network must be compared with the tracks classified by the expert user. The less the error of the network, the higher the classification performance of the neural network. However, it should be noted that this error is highly dependent on the quality of the inputs to the networks.

7. Results and discussion

Providing inputs to the neural network considering the limitations, it is then the time to train the ANN in order to perform the classification of Iranian musical tracks into different DASTGÄHs. In this section, the results of such effort are discussed.

7.1. Results for a specific sample

Testing and training data matrices, which contain 20 local maxima of frequency intervals associated with the frequency spectrum of musical tracks played by NEY instrument, were given to the neural network as inputs. The ANN receives 20 inputs which are the local maxima of the spectrum of musical tracks and gives 7 Boolean outputs which are associated with the seven principal musical modal systems or DASTGÄHs. For comparison and optimization purposes, the number of hidden layer neurons are increased from 10 to 60 by the step of 10 and then, the results are obtained. By increasing the number of neurons, performance of the neural network was increased. This means that the network was trained better at the cost of a longer processing time.

As seen in Fig. 2, training was continued until the mean square error became less than a preset value defined by the software in which the performance of the ANN is optimized.

7.2. Training error

After the ANN is trained, testing and training data covariance matrices are presented by the software. The results are shown in Tables 2 and 3. In these matrices, the number of each column and row represents a specific DASTGÄH. The numbers on the main diagonal of matrix represent the number of data that are classified correctly. However, other elements of the matrix indicate the number of inputs that the neural network has incorrectly classified. According to Table 2, the training tracks are classified correctly.

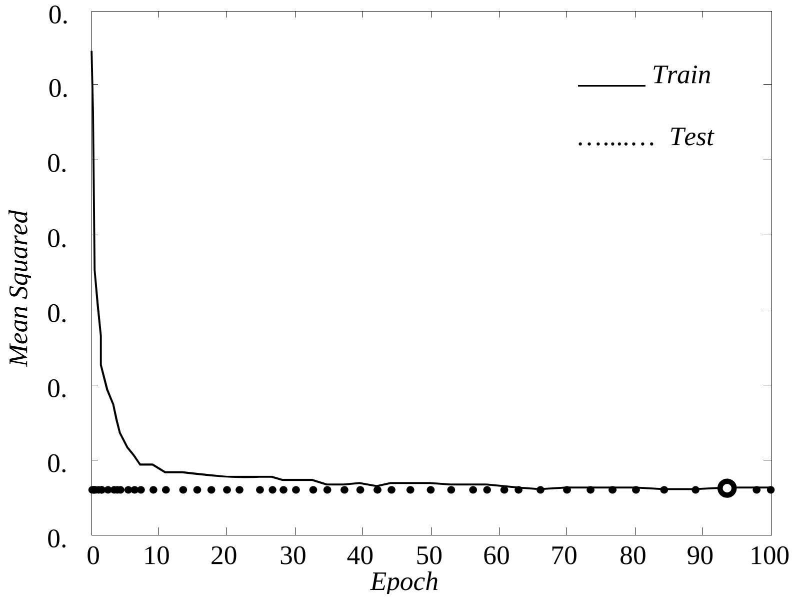


Fig. 2: Training error reduction and the stop of the process; best training performance 0.019292 at epoch 1000

Table 2. Training data covariance matrix

Class	1	2	3	4	5	6	7
1	140	0	0	0	0	0	**0**
2	0	45	0	0	0	0	**0**
3	0	0	12	0	0	0	**0**
4	0	0	0	15	0	0	**0**
5	0	0	0	0	22	0	**0**
6	0	0	0	0	0	14	**0**
7	0	0	0	0	0	0	**16**

Table 3. Testing data covariance matrix

Class	1	2	3	4	5	6	7
1	32	4	1	1	2	1	**1**
2	3	7	0	0	0	0	**0**
3	1	1	3	0	1	0	**0**
4	2	2	0	4	0	0	**0**
5	1	0	0	0	4	0	**0**
6	3	0	0	0	0	3	**0**
7	3	0	0	0	0	0	**4**

7.3. Training and test accuracy

In addition to the covariance matrices, the percentage values of data that are classified correctly in both training data and testing data are obtained. Tables 4, 5 and 6 show the classification accuracy corresponding to NEY, violin and vocal song respectively, versus the number of neurons in the hidden layer of the neural network. As can be seen, with an increase in the number of neurons, the percentage of training accuracy increases up to 100%. This means that the training data are classified correctly and the covariance matrix is diagonal. On the other hand, it is seen that by increasing the number of neurons, the accuracy of test data increases; however, when the accuracy of training data reaches 100%, then the accuracy of testing data begins to decrease. It is because when the number of neurons is much higher than the number of inputs, the number of unknowns of the problem increases and thus makes it difficult to solve the problem. Therefore, the number of hidden layer neurons should be close to the number of inputs of the network. In the experiments, it was observed that the highest percentage of correct classification (best result for accuracy of the test) for NEY in the worst case is about 65%. The case is when the pieces of music are from all GÖSHEHs of each DASTGÄH (such as FEILÏ, SHEKASTEH, NAGHMEH, MEHRABÄNÏ, JÄMEH-DARÄN, RÖHOL-ARVÄH, etc.) and there are no restrictions regarding the type of GÖSHEH. The result is about 72% for violin and 56% for vocal song.

Table 4. Training and testing accuracy changes versus the number of neurons for NEY

Neurons of hidden layer	Classification accuracy of training data	Classification accuracy of test data
10	87.197	56.8072
20	98.803	60.3735
30	100	64.9639
40	100	53.9398
50	100	39.3012
60	100	28.8453

Table 5. Training and testing accuracy changes versus number of neurons for violin

Neurons of hidden layer	Classification accuracy of training data	Classification accuracy of test data
10	95.9326	61.8043
20	100	67.2844
30	100	72.2255
40	100	60.8453
50	100	52.1209
60	100	43.2320

Table 6. Training and testing accuracy changes versus the number of neurons for vocal song

Neurons of hidden layer	Classification accuracy of training data	Classification accuracy of test data
10	82.2838	43.7602
20	92.1004	51.2043
30	100	55.8752
40	100	46.8215
50	100	37.4006
60	100	31.9420

7.4. Discussion

The results of this study show that it is possible to classify and categorize the Iranian pieces of music and distinguish their corresponding DASTGÄHs by an accuracy rate of 55 to 75 percent according to their frequency spectrum and with the aid of the Fast Fourier Transform. Although these values are acceptable but they are not completely satisfactory. It is because in the frequency spectrum, time is eliminated and so the sequence of played notes is not considered as a feature to classify the pieces of music. This error could adversely affect the results of the study which are mainly based on the analysis of the frequency content of the pieces of music.

7.5. Comparison with other related works

As mentioned in Section 1, there are some other researches about classifying the traditional music among witch there are few works who have used an approach similar to the current paper. In a similar research in the field of classifying the Iranian musical DASTGÄHs [10], serious restrictions and limitations have been considered on the input data. The experiment is performed only for one musical instrument and the pieces of music are played in DARÄMAD (the first GÖSHEH) of the MÄHÖR. Therefore, the task of the neural network is to classify the MÄHÖR from other DASTGÄHs. Given these conditions, the accuracy of the test data is about 70% which is less than some results of the present study. However, none of these restrictions are imposed to the current study and the classification is performed for all DASTGÄHs as well as for three different musical instruments (one of which is a vocal song).

8. Conclusion

In this paper we have tried to classify and distinguish the DASTGÄHs corresponding to each played piece of traditional music as much as possible in an automatic manner. In this direction, mathematical methods and computational tools have been utilized (neural networks).

To provide inputs to the neural network, pieces of music played by NEY and violin as well as vocal songs were studied. Pieces of music related to these three categories are recorded digitally in the format of "wav" and then, the frequency spectrum of these data were obtained using the Fast Fourier Transform. The resulted data were normalized in the range of 1 and -1 and the frequency range is divided into 20 intervals for comparison purposes. The maximum value of each interval was nominated as the representative of that interval.

The artificial neural network structure used in this study was a multilayered perceptron with supervised learning carried out by backpropagation. The configuration consisted of input and output layers with one hidden layer. In order to train the network, 70% of the data were used to train while 30% of them were selected as test data. The results showed that with the analysis of frequency spectrum, the accuracy of the network in worst cases was more than 64% for NEY, more than 72% for violin and more than 55% for vocal song.

Obviously, if we set limitations and restriction on the experiment conditions and preparation of the input data, better results would be achieved. It is also important that FFT eliminates the time factor from data resulting in confusing the network in distinguishing the different DASTGÄHs of the played pieces of music. Some other approaches like wavelet may help in this regard.

References

[1] P. Scott, Music classification using neural networks, Manuscript Class ee373a, Stanford, (2001).
[2] J.-W. Lee, S.-B. Park, S.-K. Kim, Music genre classification using a time-delay neural network, in: Advances in Neural Networks-ISNN 2006, Springer, 2006, pp. 178-187.
[3] Z. Cataltepe, Y. Yaslan, A. Sonmez, Music genre classification using MIDI and audio features, EURASIP Journal on Advances in Signal Processing, 2007 (2007) 1-8.
[4] K. Kosina, Music genre recognition, (2002).
[5] H. Habibi, M. HomayoonPour, Automatic detection of music styles, Signal and Data Processing, 1 (2010) 33-52.
[6] N. Darabi, N. Azimi, H. Nojumi, Recognition of Dastgah and Maqam for Persian music with detecting skeletal melodic models, in: Proc. 2nd IEEE BENELUX/DSP Valley Signal Processing Symposium, Citeseer, 2006.
[7] S. Abdoli, Iranian traditional music Dastgah classification, in: ISMIR, 2011, pp. 275-280.
[8] M.A. Layegh, S. Haghipour, Y.N. Sarem, Classification of the Radif of Mirza Abdollah a canonic repertoire of Persian music using SVM method, Gazi University Journal of Science Part A: Engineering and Innovation, 1 (2013) 57-66.
[9] H. Hajimolahoseini, R. Amirfattahi, M. Zekri, Real-time classification of Persian musical Dastgahs using artificial neural network, in: 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP), IEEE, 2012, pp. 157-160.
[10] S. Mahmoodan, A. Banooshi, Automatic classification of Iranian music by an artificial neural network, in: 2nd International Conference on Acoustics and Vibration, 2012.

[11] N. Darabi, Generation and analyzation digital signals of music and automatic recognition of music styles, in, M.Sc. Thesis, K. N. Toosi University of Technology, 2004.
[12] H. Farhat, The Dastgah concept in Persian music, Cambridge University Press, 2004.
[13] C.J. Plack, A.J. Oxenham, R.R. Fay, Pitch: neural coding and perception, Springer Science & Business Media, 2006.
[14] J.G. Roederer, The physics and psychophysics of music: an introduction, Springer Science & Business Media, 2008.
[15] D.A. Ross, Being in time to the music, Cambridge Scholars Publishing, 2008.
[16] R. Lyon, S. Shamma, Auditory representations of timbre and pitch, in: Auditory computation, Springer, 1996, pp. 221-270.
[17] B. Yegnanarayana, Artificial neural networks, PHI Learning Pvt. Ltd., 2009.
[18] G. Tzanetakis, P. Cook, Musical genre classification of audio signals, IEEE transactions on Speech and Audio Processing, 10 (2002) 293-302.