International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 34813-34819

A Statistical Study of Artist Style from Raga Performance

Poonam Priyadarshini

Deptt. of Electronics and Communication Engineering Birla Institute of Technology, Mesra Patna, India
ppbitpatna@gmail.com

Soubhik Chakraborty

Department of Mathematics Birla Institute of Technology, Mesra Ranchi, India
soubhikc@yahoo.co.in

Abstract
Music Information Retrieval (MIR) is an emerging research area that receives growing attention from both the research community and music industry. It addresses the problem of querying and retrieving certain types of music from large music database. People searched music of their choice from database by song title, composer and performer. Features of interest may include melody, harmony, rhythm, and instrumentation.
Although every raga has a characteristic mood of its own, differences occur between renditions of the same raga based on the gharana or personal style of the artist.This paper discusses the use of projection pursuit techniques along with MFCC to analyze the similarity between different artists singing the same raga.

Keywords— MFCC, projection pursuit, raga, Gharana, artist classification.

I. INTRODUCTION

Though Indian Classical is also considered to be a major form of music, the current literature related to it is very limited, in comparison to its western counterpart. Indian classical music is known for its technical soundness and its well defined structure. In this paper we address the analysis of Indian music; in specific, we attempt to develop techniques to study the various features of ragas. According to the music theory, musical signal can be analyzed from the aspect of dynamics, rhythm, timbre, pitch and tonality [4].

A raga is a melodic piece of music which is characterised by the underlying units of sounds called swaras. Raga is a collection of different unique notes that are having some special properties (e.g. aroha, avaroha, Taal, Pakad, etc.). Each Raga consists of unique sequential arrangement of swara or notes. Notes are called swaras in Indian classical music. The basic seven notes or swaras or symbols in classical music are S (Sa),R (Re or Ri), G (Ga), M(Ma), P (Pa), D (Dha), N (Ni) which can be considered as analogous to the C, D, E, F, G, B, A[1,2] of western music. In another way we have 12 swaras or shrutis in Carnatic music, S, r, R, g, G, rn, M, p, d, D, n and N. Moreover, Indian classical music allows the artist to improvise over the definitions of a raga to create his own personal performance of the raga. Due to this, two performances of the same raga by different artists may sound strikingly different to the novice ears, though they still maintain the defining qualities.Since MFCC is a powerful and popular signal processing tool that can provide a clearer scientific picture of the musical pieces rather than provide just a set of acoustical features, we use MFCC to get a multidimensional point that represents the signature of the artist scientifically and then convert it to two dimension,whereupon it becomes a curve, using projection pursuit. The literature on projection pursuit suggests that if two multidimensional points are close, so would be the corresponding two dimensional curves. The rest of the paper is organized as follows: Section II describes the relevant literature work for recognition of ragas. In section III, the proposed method is based projection pursuit techniques along with MFCC to analyze the similarity between different artists singing the same raga. Section IV provides analysis of the proposed method and result respectively.

II. RELATED WORK

A large number of distinct feature sets, mostly developed for speech recognition, have been proposed to represent audio signals. Typically Audio signals are based on some form of time-frequency representation. Mel-frequency cepstral coefficients (MFCC) [2], illustrates the shape of the spectrum of the audio signal and are widely used in speech recognition [2]. In Indian classical music mainly the research performed on automatic Indian music information, recognition, and classification of *ragas* and development of retrieval systems [18]. Hariharan S. et. al [4] used Hidden Markov Model and Neural Network for recognition of *ragas* and notes with use of *Arohana* and *Awarohana* sequence. Gaurav Pandey [5] suggested note transcription system 'TANSEN' and used Hidden Markov Model and string matching technique to get *pakad* notes of *ragas*. Chordia et. al [6] used recognition of annotation, onsets detection, pitch class distribution (PCD) and pitch class dyad distribution (PCCD) using SVM, MVN and Random Forests. In this Chordia also attempted classification of tone profiles and spectral profile, pitch detection, onset detection, PCDs and PCCDs using HMM, MVN, FFNN,KNN,Tree based and Bayesian classifier. Neural Network self organized Maps (SOM) and Bayesian decision rule for recognition of pitch profile, pitch class distribution are described in [3] [4]. Sridhar R. and Geetha T.V. [9][10] used signal separation, segmentation and string

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 34813-34819

matching for signal frequency offset and onset. In [11] statistical framework is used for vocal pitch, folded pitch distribution (FPDs), *Swara* annotation and pitch class distribution (PCDs). Chordia et. Al [12] used MFCC, GMM and PCD for retrieval of timbre and PCD for artist recognition, instrument recognition and *Thaat* classification. Shettty S.[13] used ANN for note transcription and *Arohana*, *Awarohana* pattern for *raga* recognition.

III. FEATURE EXTRACTION

Feature extraction is a crucial step to improve the accuracy of the classification. In common analysis, the musical features are always divided into three categories according to the procedure of the signal processing, i.e., the feature of time domain, the frequency domain and the time-frequency domain. Here, the feature of time domain involves numerical features by directly analyzing the sound wave, including short-time energy (STE), short time zero cross rate (ZCR), liner predictive coefficient (LPC) [2].The features in frequency domain mainly reflect the information in frequency by using Fast Fourier Transform (FFT). A series of further features are proposed such as frequency energy and Mel frequency ceptral coefficients (MFCC). The time-frequency feature mainly refers to the further calculation based on the wavelet transformation, which could reflect the characters of the signal in time and frequency domain simultaneously [9].

The MFCC is an important auditory feature which reflects the hearing characteristics of human being, and it has been widely used in the audio processing. MFCC works by producing a MFCC matrix. First, by using fast Fourier transformation (FFT), the signal is transformed into the Mel scale. Second, the Mel scale triangular filter banks are applied to the signal, after that, the discrete cosine transformation (DCT) is performed, and then MFCC is obtained for each frame.The other features in the timbre consist of the statisticfeature and some easy calculate features such as zero cross and brightness of the signal. Here the zero cross means the number of times that the signal crosses the x-axis. Other statistic features including the centroid,spread, skewness, kurtosis and flatness, could be referred to[1].

IV. MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC)

The use of Mel Frequency Cepstral Coefficients can be considered as one of the standard method for feature extraction (Motlíček, 2002). The use of about 20 MFCC coefficients is common in Music Analysis, although 10-12 coefficients are often considered to be sufficient for coding music (Hagen at al., 2003). The most notable downside of using MFCC is its sensitivity to noise due to its dependence on the spectral form. Methods that utilize information in the periodicity of music signals could be used to overcome this problem The non-linear frequency scale used an approximation to the Mel-frequency scale which is approximately linear for frequencies below 1 kHz and logarithmic for frequencies above 1 kHz. This is motivated by the fact that the human auditory system becomes less frequency selective as frequency increases above 1 kHz. The MFCC features correspond to the cepstrum of the log filterbank energies.

MFCC Coefficients of the samples are as follows:

Table 1: MFCC value for Gwalior Gharana

GWALIOR GHARANA				
RAGA	**Todi**	**Bihag**	**Todi**	**Bihag**
ARTISTS	**Telang**	**Pandit**	**J. Iyer**	**J.Iyer**
MFCC1	**3.9642**	**4.4889**	**7.3051**	**2.8566**
MFCC2	**1.1885**	**0.6179**	**1.3231**	**1.4701**
MFCC3	**0.3436**	**0.3375**	**1.8681**	**0.5350**
MFCC4	**0.3265**	**0.2048**	**0.8152**	**0.2852**
MFCC5	**0.2280**	**0.0990**	**0.4928**	**0.2376**
MFCC6	**0.2799**	**0.1367**	**0.4207**	**0.2856**
MFCC7	**0.2717**	**0.1479**	**0.4213**	**0.2315**
MFCC8	**0.6053**	**0.0694**	**0.5231**	**0.2047**
MFCC9	**0.2912**	**0.1804**	**0.2496**	**0.1586**
MFCC10	**0.0963**	**0.3914**	**0.2390**	**0.1613**
MFCC11	**0.1282**	**0.2108**	**0.2046**	**0.1272**
MFCC12	**0.1515**	**0.0839**	**0.1735**	**0.2559**
MFCC13	**0.0955**	**0.0818**	**0.2427**	**0.1645**

Table 2: MFCC coefficient of Kirana Gharana

KIRANA GHARANA			
RAGA	Todi raga	Bihag raga	Bihag raga
Artists	A. Ali	S.Bhattacharjee	Khan
mfcc1	4.2345	4.4873	3.5750
mfcc2	1.1334	1.5249	3.2114
mfcc3	0.1759	0.4286	0.3437
mfcc4	0.3569	0.3541	0.3335
mfcc5	0.2701	0.2124	0.4561
mfcc6	0.2122	0.1659	0.3172
mfcc7	0.2933	0.2611	0.1581
mfcc8	0.3332	0.1769	0.1942
mfcc9	0.1548	0.1649	0.3062
mfcc10	0.1492	0.1876	0.2003
mfcc11	0.0991	0.0963	0.2041
mfcc12	0.1385	0.1317	0.3464
mfcc13	0.1279	0.1341	0.1569

0.156

Statistical parametrization:

The designed Statistical parametrization approach is aimed at describing structural feature of musical phrase. The Statistical parametric description has been used in previous case studies (Kostek, 2005),and introduced parameter are as follows.

P1 = The difference between weighted average note pitch and the pitch of the lowest note of a phrase.

$$P_1 = \frac{\sum_{i=1}^{N} p_i d_i}{T} - M_{in}(p_i)$$

where T is the phrase duration,denotes the pitch (at the onset) of the i-th note and denotes the duration of the i-th note(departure of the i-th note- onset of the i-th note), N denotes the number of note in a phrase.[19]

P2 = The difference between the pitch of the highest and the lowest note of a phrase

$$P_2 = Max(p_i) - Min(p_i)$$

P3 = The average absolute difference of the pitches of subsequent note.

$$P_3 = \frac{1}{N-1}\sum_{i=1}^{N-1}|p_i - p_{i+1}|$$

P4 = The duration of the longest note of a phrase.

$$P_4 = Max(d_i)$$

P5 = The average note duration.

$$P_5 = \frac{1}{N}\sum_{i=1}^{N} d_i$$

F (*Mel*) =[2595*log 10[1+ *f*]700]

The first few coefficients obtained by this method contain most of the signal energy. Hence, usually the first five coefficients are taken as features.

Using MFCC a feature set was created consisting of statistical and acoustic features to analyze the audio samples. There are 12 features in all leading to a 12 dimensional point to represent the artist's signature or identity.

B. Statistical Features

1. *Skewness :* Skewness is the measure of the symmetry or asymmetry of a signal.

$$Skewness = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^3}{(N-1)s^3}$$

Y is the mean, *s* is the standard deviation, and *N* is the number of data points.

2. *Kurtosis:* Kurtosis is a measure of the noisiness of a signal. It shows whether the data is peaked or flat relative to a normal distribution.

$$Kurtosis = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})4}{(N-1)s^4}$$

Projection pursuit methods are very useful tools in the area of high-dimensional data analysis. They do however present many limitations as enumerated by Crawford and Fall [17]. The technique employed here for dimensional reduction is the Andrews plot method. An Andrews plot is a method to visualize multi-dimensional data in two dimensions. Each point x = {x1, x2,…,xn}in the dataset denotes a finite Fourier series.

$$f_{x(t)} = \frac{x_1}{\sqrt{2}} + x_2\sin(t) + x_3\cos(t) + x_4\sin(2t) + x_5\cos(2t) + \cdots ....$$

This function is then plotted for $-\pi < t < \pi$.

Each point is now a line between $-\pi$ and π. Any structure in the dataset is now visible as curves on a two-dimensional plot. By a property of Andrews plots, any if any two N-dimensional points are close, the corresponding curves in two dimensions are also close accordingly. Hence, we can identify clusters among the curves and tell which artists are close in rendering the same raga. If two artists train under the same teacher, by comparing the performances of the two artists with that of the teacher's we can determine which artist more closely resembles the teacher's individual style. Artists from the same gharana, or school tought, are more likely to produce curves closer to one another when they render the same raga. For artists from different Gharanas, the curves will be further apart. Thus projection pursuit can help in the classification of renderings of ragas according to the Gharana of the artist responsible for the particular rendition.

V. ANALYSIS AND RESULTS

We are choosing Hindustani (North Indian) raga eg. Todi Raga and Bihag Raga for doing analysis of music signal.

Statistical parameters representing a musical phrase can be divided into two groups: parameter describing melodic quantities of musical phrase $(P_1, P_2 P_3)$ and the parameter describing rhythmical quantities of musical phrase (P_4, P_5). we are finding the pitch and time duration between notes from praat software, the value of $(P_1, P_2 P_3)$ (P_4, P_5) are calculated by C++ programming then through Matlab simulation Andrew plot is created. when artists from the same Gharana are renderings the same raga then there is closeness in the curve and there is separation in the curve when the same raga is rendered by artists of different Gharanas. The result are shown in the figures below of feature extracted by MFCC of todi segment 1 and 2 and the Andrews plots of Todi Raga and Bihag Raga sung by various Artists.

The values p1=1006.018790 p2 =108.371178 p3 = 26.970213 p4 = 7.437506 p5 = 7.496207 of Todi segment1 denoted by (F1) and the values of Todi segment 2 denoted by(F2) is p1 = 445.011141 p2 = 41.764107 p3 = 13.241155 p4 = 4.787417 p5 = 5.895031 and similarly other values are calculated.

Table 3 : Values of statical parametrization of Todi and Bihag raga of Gwalior Gharana

GWALIOR GHARANA				
	TODI RAGA(1)	BIHAG RAGA(1	TODI RAGA(2)	BIHAG RAGA(2)
ARTISTS	B. Telang	M. Pandit	J. Iyer	J. Iyer
P1	52.7220	371.781	96.1620	10.8434
P2	90.6974	149.920	145.764	103.323
P3	18.9497	30.5496	93.9135	23.4139
P4	2.0950	2.9150	2.6160	1.0780
P5	0.8685	1.7854	0.09873	0.5183

Table4: Values of statical parametrization of Todi and Bihag raga of Kirana Gharana

KIRANA GHARANA			
	TODI RAGA(1)	BIHAG RAGA(1)	BIHAG RAGA(2)
ARTISTS	A. Ali	S. Bhattacharjee	S. Khan
P1	286.605	275.890	142.503
P2	157.467	112.585	183.852
P3	86.3762	44.9000	72.9682
P4	3.59600	5.63700	2.05600
P5	2.14250	2.21776	1.01828

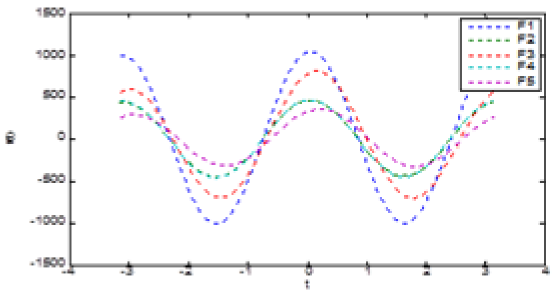

Figure 1. Andrews plots of Todi Raga by various Artists

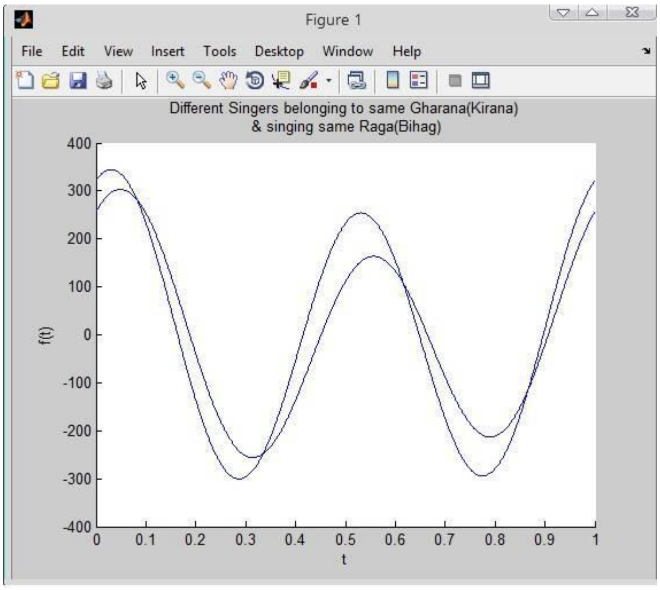

Figure 2 Andrews plot for different singer of same Gharana, same raga

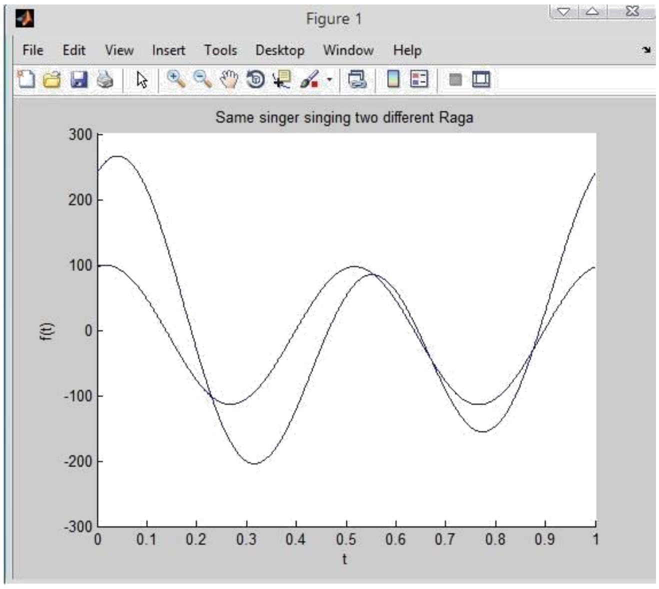

Figure 3. Andrews plot for same singer singing different raga

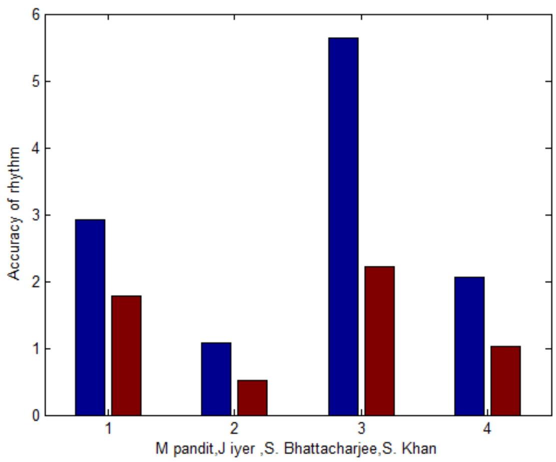

Figure 4. Rhythm plot for different singer singing same raga

Music classification is crucial for the categorization of bulky amount of music content. Automatic music classification finds important applications in professional media production, radio stations, audio-visual archive management, entertainment and others. Although it is hard to precisely define the specific content of a music ragas, it is generally agreed that audio signals of music belonging to the same ragas contain certain common characteristics since they are composed of similar types of instruments and having similar rhythmic patterns. These common characteristics motivated recent research activities to improve automatic music genre classification.
For example we have taken all ragas (Todi and Bihag) MFCC of kirana gharana as target and the one todi by Arshad ali of kirana gharana as input to the network.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 34813-34819

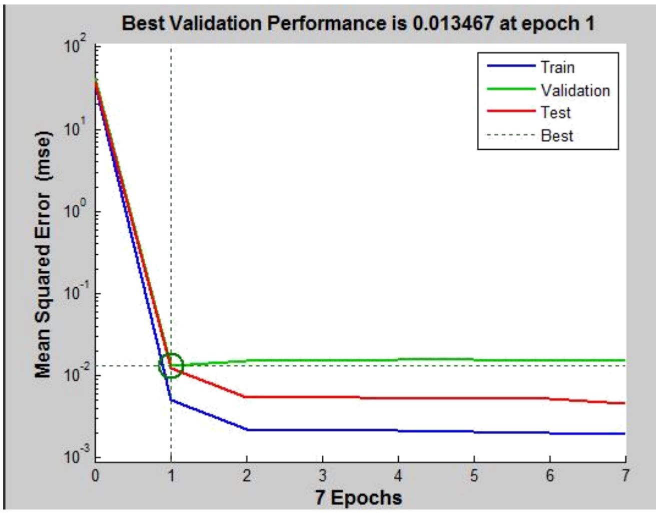

Figure5.Network Performance

MSE is a network performance function. It measures the network's performance according to the mean of squared errors.If the testing and validation graph are more along and alike, and the testing plot lies below the validation then we can conclude that classification is more accurate.

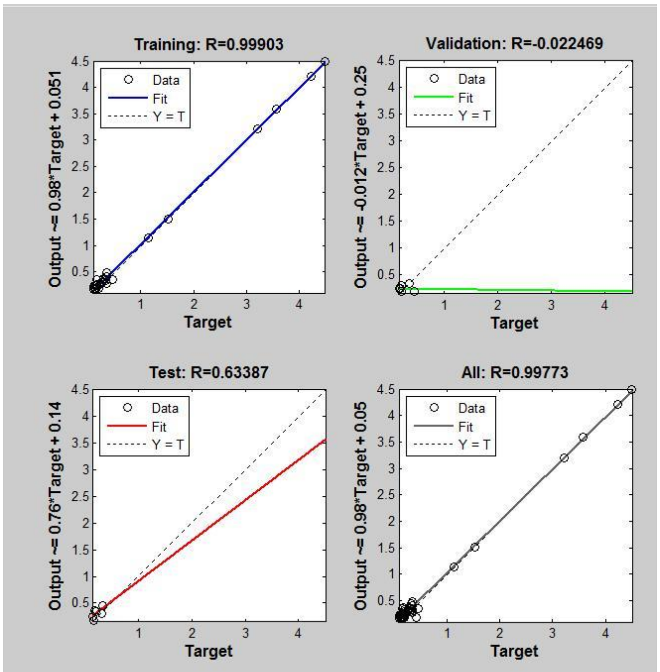

Figure6.Output versus target during testing, training, validation

CONFUSION MATRIX

A confusion matrix, also known as a contingency table or an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix). Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class.[19]

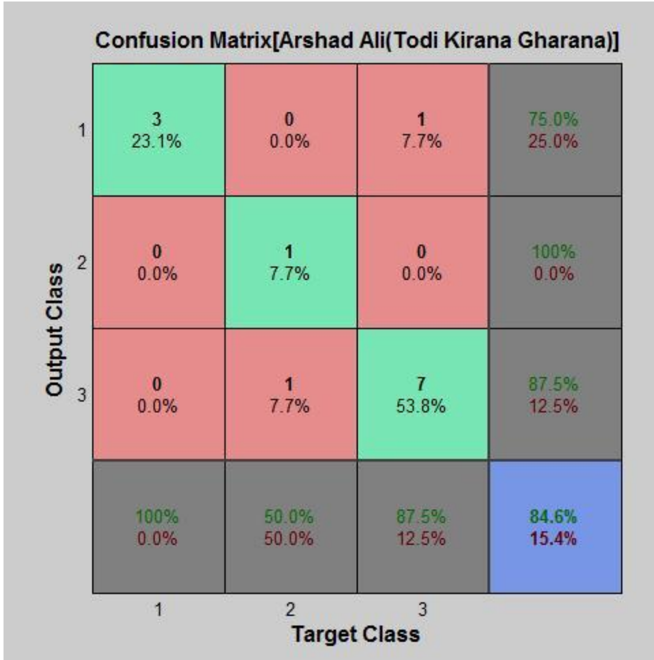

Figure 7.Confusion Matrix Todi raga

Another example Gwalior gharana having segment of two Todi raga and two Bihag raga MFCC as target for network and one bihag raga as input.

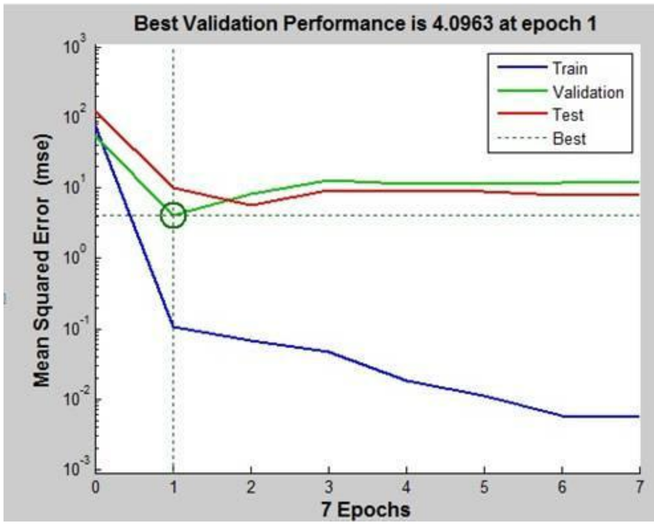

Figure8.Network performance for Gwalior gharana

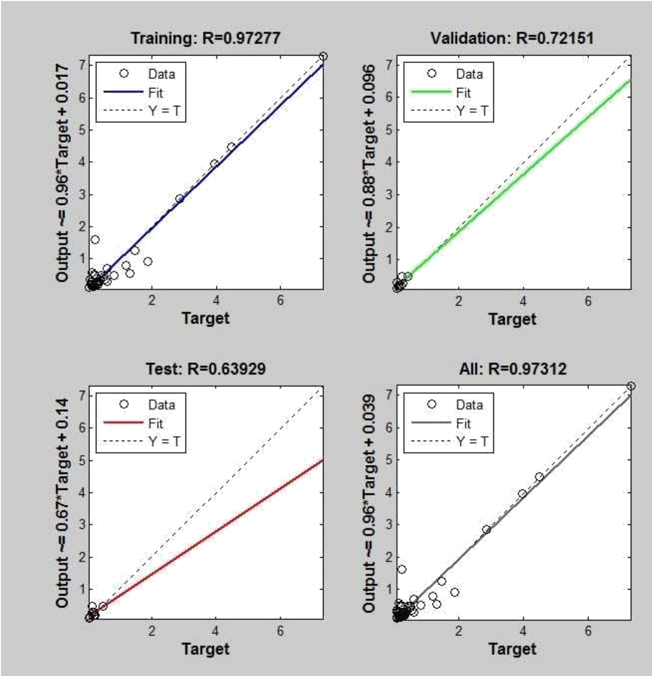

Figure 9.Output versus target during testing, training, validation

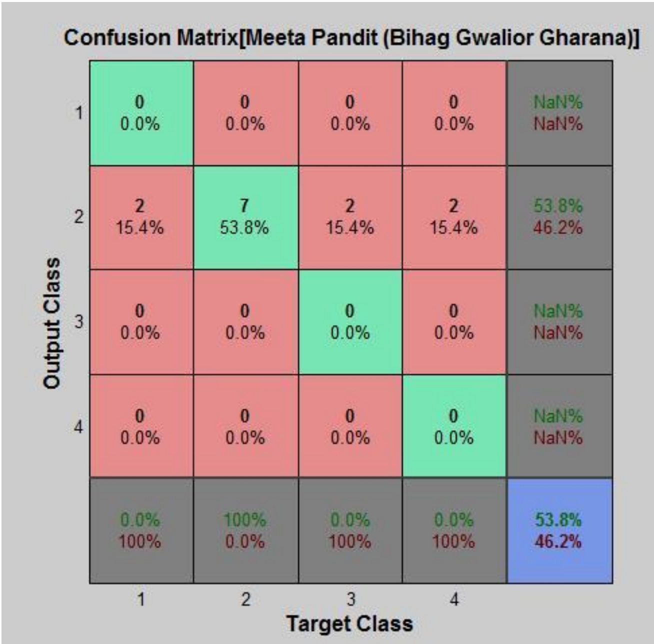

Figure 10.Confusion Matrix Bihag raga

VI. CONCLUSION

From the Andrews plots of todi raga by various Artists we saw that closeness in the curve is more at f(t)=100 at t=-.8 to +.8 and t=-2.2 to +2 and there is separation in the curve at f(t)=1000 at t= -1.5 to +1.5. Todi segment1denoted by F1 is showing more separation in the curve so this artists belong to different Gharanas where as F4 and F5 shows closeness in the curve so these artists belong to same Gharanas. Music style classification is one of the key problems in MIR.

In this paper for feature extraction MFCC and LPC were used and for classification Artificial Neural Network was used. The performance of the training and test datasets are averaged to obtain the overall percent efficiency of the system. The percentage efficiency is defined as the ratio of correct classifications to the total test cases. It is clear that as the number of inputs or features is increased, the performance of the system increases. The performance is also enhanced when the duration of the test piece is increased.

REFERENCES

[1] J.P.Belloand J. Pickens, “A robust mid-level epresentation for harmonic content in music signals,” in *Proc. ISMIR*, pp. 304–311, 2005.

[2] L. R. Rabiner and B.H.Juang- Fundamentals of speech recognition; Prentice Hall,New Jersey, 1993.

[3] P.Dighe and H.karnick,"Swara histogram based structural analysis and identification of Indian classical music", in ISMIR,2012.

[4] Koduri,Gopala Krishna,Sankalp Gulati, and Preeti Rao. "A survey of raga recognition techniques and improvements to the state-of-the-art."*Sound and Music Computing* (2011).

[5] Pandey G., Mishra C., and Paul I., “Tansen- A system for automatic raga identification”, In Proc. of the 1st Indian International Conference on Artificial Intelligence, 2003.

[6] Chordia P., and Alex Rae, “Automatic Raga classification using pitch-class and pitch-class dyad distributions”, I Proc. of the 7th International Conference on Music Information Retrieval(ISMIR), 2007.

[7] ChordiaP.,and Rae A.“Modeling and visualizing tonality in North Indian classical music”, In Neural Information Processing Systems(NIPS),Music Brain Workshop, 2007.

[8] ChordiaP.and RaeA.,“Raag vidya- real-time raag recognition for interactive music”, In Proc.of the International Conference on New Interfaces for Musical Expression (NIME), Genova, Italy, 2008.

[9] Sridhar R. and Geetha T.V., “Swara identification for South Indian Classical Music”, 9th International Conference on Information Technology (ICIT'06), pp 143-144, 2006.

[10] B. Kostek, Perception based data processing in acoustics: Applications to Music Information Retrieval and Psychophysiology of hearing, Springer, 2005

[11] Belle S., Joshi R. and Rao P.“Raga identification by using Swara intonation”, Journal of ITC Sangeet Research Academy, vol. 23, 2009.

[12] Chordia P., GodfreyM. and Alex Rae, “Extending content-based recommendation- The case of Indian classical music”, Content-Based Retrieval, Categorization and Similarity, ISMIR 2008.

[13] Shetty S. and Achary K.K,“Raga mining of Indian Music by extracting Arohana-Avarohana pattern”,

International Journal of Recent Trends in Engineering Vol. 1, No. 1, 2009.

[14] P.Chordia and ARae,"Raag recognition using Pitch-class and pitch-class dyad distributions", in Proc. of ISMIR,2007,pp. 431-436.

[15] R.Sridhar and T.Geetha,"Raga identification of carnatic Music for music information retrieval",in International Journal of Recent Trends in Engineering vol.l,no.I, 2009,pp.571-574

[16] K. Mahto, A. Hota, S. S. Solanki and S. Chakraborty, A Study on Artist Similarity Using Projection Pursuit and MFCC: Identification of Gharana from Raga Performance, Proceedings of the 8th IndiaCom 2014, 2014 International Conference on "Computing for Sustainable Global Development", BVICAM, New Delhi, 5-7 March, 2014 (included in IEEE Explore; 978-93-80544-12-0/14; p.647-653)

[17] Crawford, Stuart L. and Fall, Thomas C.,“Projection pursuit techniques for visualizing high-dimensional data sets”, Visualization in Scientific Computing, (G. M.Nielson and B. Shriver, eds.), IEEE Computer Society Press,pp.94-108,1990.

[18] Trisiladevi C. Nagavi Nagappa U. Bhajantri ‘Overview of automatic Indian music information recognition, classification and retrieval systems”2011 International Conference on recent Trends Information Systems.

[19] Changsheng Xiang, ZiYing Zhou, “A New Music Classification Method based on BP Neural Network International Journal of Digital Content Technology and its Applications.” In Volume 5, Number 6, June 2011.

[20] Nidhi Srivastava, “Speech Recognition using Artificial Neural Network” in International Journal of EngineeringScience and Innovative Technology (IJESIT) Volume 3, Issue 3, May 2014.