

Hybrid Feature Based Classifier Performance Evaluation of Monophonic and Polyphonic Indian Classical Instruments Recognition

Abhijit V. Chitre¹, Ketan J. Raut², Tushar Jadhav³, Minal S. Deshmukh⁴
and Kirti Wanjale⁵

^{1,2,3,4} Dept. of E&TC Engineering, Vishwakarma Institute of Information Technology,
Pune, India

⁵ Dept. of Computer Engineering, Vishwakarma Institute of Information Technology,
Pune, India

¹ abhijit.chitre@viit.ac.in, ² ketan.raut@viit.ac.in, ³ tushar.jadhav@viit.ac.in,
⁴ minal.deshmukh@viit.ac.in, ⁵ kirti.wanjale@viit.ac.in

Abstract: Instrument recognition in computer music is an important research area that deals with sound modelling. Musical sounds comprises of five prominent constituents which are Pitch, timber, loudness, duration, and spatialization. The tonal sound is function of all these components playing critical role in deciding quality. The first four parameters can be modified, but timbre remains a challenge [6]. Then, inevitably, timbre became the focus of this piece. It is a sound quality that distinguishes one musical instrument from another, regardless of pitch or volume, and it is critical. Monophonic and polyphonic recordings of musical instruments can be identified using this method.

To evaluate the proposed approach, three Indian instruments were experimented to generate training data set. Flutes, harmoniums, and sitars are among the instruments used. Indian musical instruments classify sounds using statistical and spectral parameters. The hybrid features from different domains extracting important characteristics from musical sounds are extracted. An Indian Musical Instrument SVM and GMM classifier demonstrate their ability to classify accurately. Using monophonic sounds, SVM and Polyphonic produce an average accuracy of 89.88% and 91.10%, respectively. According to the results of the experiments, GMM outperforms SVM in monophonic recordings by a factor of 96.33 and polyphonic recordings by a factor of 93.33.

Keywords: Indian Classical Instruments, Support Vector Machine (SVM), Gaussian Mixture Modelling, k-nearest neighbor classifier (KNN)

¹ Corresponding Author

1. Introduction

Musical instrument identification exhibits number of applications which are useful in musical data mining. Using such identification system we can directly categorize or search in musical audio database without need of descriptive data included in audio signal. Musical instrument identification is a key aspect of music information retrieval task. If we know instruments involved in a given musical piece, we can explore very useful information. The applications of musical content analysis are structural coding, automated musical signal annotation and musician's tool etc. Many other research fields are also related to musical instrument identification system. The paper proposes a classification mechanism for features of multiple domains and presents the accuracy results of classifiers. The overall system is shown in the Figure 1.

Indian music is one of the old traditions in the world. Musical instruments are the basis of Indian music. Most of the Indian musical instruments have developed gradually. Different types of instruments are present in Indian music such as stringed instruments, percussion instruments and wind-blown instruments. We have considered three most commonly used Indian classical instruments i.e. flute, harmonium and sitar. Music information retrieval system reveals the classic application of instrument classification and automatic indexing [1]. Automatic musical instrument classification demonstrates a crucial function in facilitating the rapid discovery and organization of musical data. Musical data analysis has practical applications in a variety of areas, including database retrieval systems and automatic musical signal annotation etc. [2][13].

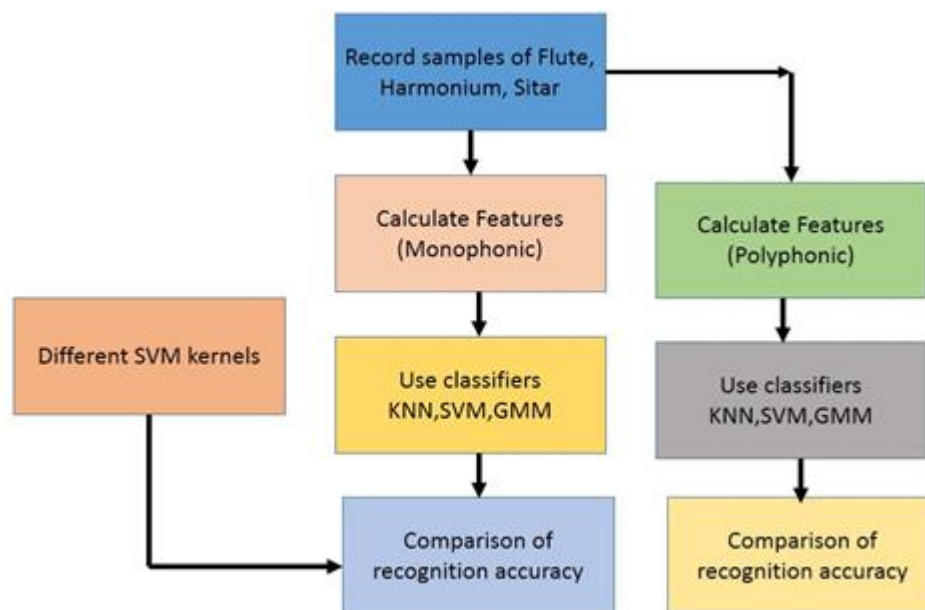


Figure 1. Block Diagram of the Proposed System

Feature extraction is a very important part of any identification system. Feature extraction process starts from input data or preprocessed data which is large, redundant and transformed into reduced data i.e. features which are informative, non-redundant [3]. It is expected that the extracted features must contain relevant information from the input data. These features are used for classification purpose instead of complete data. Different feature extraction algorithms with required mathematics are described here. The features employed in this work are derived from previous studies in speech and have been found to be beneficial for identifying western musical instruments.

Following are the different parameter analysis methods used in this work:

1. Statistical features
2. Temporal and spectral features.

2. Literature Survey

Substantial research has been reported in literature to explore different methods for automatic recognition of musical instrument. The major research can be found on Western Musical instrument such as piano, clarinet, Trumpet, Guitar, Violin. As stated before, this work focusses predominantly recognition of Indian Musical Instruments such as Flute, Harmonium, Sitar in Monophonic and Polyphonic Recordings.

Daniel Mintz's work [5] describes a revolutionary technique to sound synthesis based on timbre. The proposed approach for characterising timbre quality is strongly influenced by spectral shape, amplitude envelope, and spectro-temporal fluctuation, and the timbral separation transforms these parameter values into direct envelopes.

Juan José Burred et al. [1] discussed a framework for encoding timbre models, which was critical for instrument sample classification. The suggested work consists of a spectrotemporal envelope with unknown sample PCA and a classifier based on GMM. This method claims to have a classification accuracy of 94.9% when used with a five-class database. The technique provides precise timbre modelling using several descriptors. Spectral, temporal, spectro-temporal, and statistical parameters are extracted.

Swe Zin Kalayar Khine et al. [10] discussed a framework for judging a singer's timbre quality. The suggested work consists of a vocal detection algorithm that utilises vocal parameter extraction from each song and a cepstral coefficient, both of which are critical for vocalist timbre recognition. GMM is used to classify this parameter. Singer detection had a typical error rate of 12.2%, according to the classification. Poli [11] devised a signal processing-based approach for calculating the spectral envelope of MFCC data that can be used to identify singers. The algorithm's accuracy was determined by comparing manually detected and automatically detected sound samples.

Sang Hyun Park et al. [2] offer a technique for identifying musical instruments via timbre classification. The computing of feature vectors is the most crucial component of any automatic musical classification or analysis system. While various classifiers have been studied and compared [4], the selection of features has a greater effect on recognition accuracy than the classifiers chosen. Numerous characteristics can be utilised to characterise audio signals. They are often classified into spectral and statistical domains. The algorithm use machine learning to extract a musical instrument from complex music using timbre classification techniques such as KNN, which achieves an accuracy of 75.12%. Marques et al. [8] pioneered a system for robust musical instrument detection. The proposed method makes use of a variety of features including SVM and GMM classifiers.

From the literature review it is evident that polyphonic and monophonic hybrid feature based classification can be explored as a new paradigm to come up with the classifier and feature pair to extract maximum recognition accuracy

3. Preliminaries

Raw data or input data have random set of values. These statistics gives constant value which helps in characterizing the population i.e. data. When set of values have a tendency to cluster around some particular values, then this is useful for characterizing major of central tendency. Some statistics helps in knowing graphical shape of population. The Statistical features considered for our work are Mean, Variance, Standard deviation, Skewness, Kurtosis, Average power and Average Magnitude.

3.1 Temporal and Spectral Features

Temporal features represent time domain features which exhibit simplicity of extraction with easy physical interpretation. The spectral features (frequency-based features) are obtained using the Fourier Transform, which converts time domain to frequency domain signals. These characteristics are also referred to as spectral form features because they are utilized to describe the spectral shapes of the signal. They are often taken from the time-frequency domain's short time spectrum.

3.2 Bayesian Classification Using Gaussian Mixture Model (GMM)

The purpose of implementing a Bayesian classifier is to ensure that it can handle any amount of data dimensions, classes, and Gaussian components that is practical. Gaussian mixture models are a weighted sum of Gaussian probability density functions, which are often referred to as Gaussian components of a mixed model [10]. In fact, training a classifier is a form of supervised learning [19]. Bayesian classification is based on the notion of selecting the most likely class or the class with the lowest risk frames the fundamental basis of Bayesian.

Consider a classification task in which a feature vector is to be classified into classes [9]. Let the feature vector is to be $f = [f_1, f_2, \dots, f_D]^T$ where D is the dimension of vector. Posteriori probabilities can be computed with the Bayes formula as:

$$P(w_k | f) = \frac{p(f|w_k)P(w_k)}{p(f)} \quad (1)$$

Where, $P(w_k | f)$ is the pdf of class w_k in the feature space and $P(w_k)$ is the a priori probability, which tells the probability of the class before measuring any features. If it is unknown, then it is estimated from proportion of class in the training set $p(f)$ emulates a scaling factor ensuring the true probabilistic nature of posterior probability set.

Choosing a class with high posterior probability gives minimum error probability. Mean μ and variance σ^2 characterize Gaussian probability density function. It can be defined as

$$\mathcal{N}(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right] \quad (2)$$

where, μ is mean vector and Σ is covariance matrix.

GMM is combination of several Gaussian distributions so different subclass inside one class can be represented by using GMM model. PDF is defined as given below [8]

$$p(x; \theta) = \sum_{c=1}^C \alpha_c \mathcal{N}(x, \mu_c, \Sigma_c) \quad (3)$$

where, α_c is weight of component c , $0 < \alpha_c < 1$ for all components. $\sum_{c=1}^C \alpha_c = 1$.

Figure 2 depicts the principal of GMM.

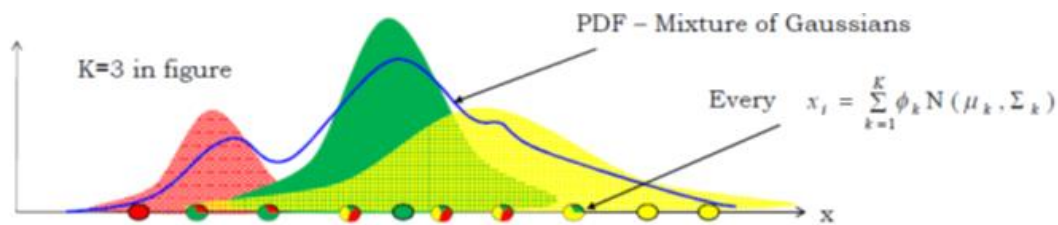


Figure 2. GMM Classification Principle

3.3 k-nearest neighbor classifier (KNN)

The purpose of KNN classifier utilizing a database of data samples of multiple classes to predict the classification of new sample [10]. KNN is one type of a distance based classifier. It is lazy learning algorithm. It stores all the training samples, and then it computes distance

between test sample and all training samples [7]. Depending upon k-nearest neighbors, the class with closest training sample will be the classification result or class with more occurrences among k-nearest neighbor will be the classification result. Different distance matrix considered for distance measurements are Euclidian, city-block, cosine, correlation and hamming. Different rules are present for how to classify the samples and they are nearest, random and consensus. The performance of KNN depends upon chosen distance matrix and the applied value of K. we can chose k as 1, 3, 5.etc. It means we can select one neighbor, three neighbors as so on for classification purpose. Figure 3 shows the clusters shown with KNN as a demonstrative picture.

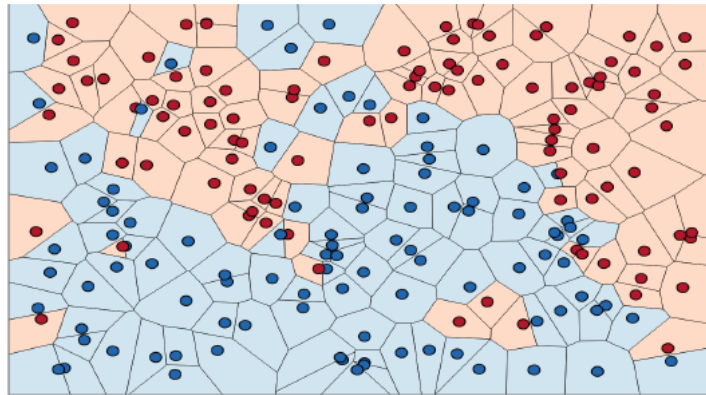


Figure 3. KNN Classification

3.4 Support Vector Machine (SVM)

SVMs are a subset of supervised learning techniques. Supervised learning is a type of machine learning method that makes predictions using a known dataset. SVMs classify objects by constructing hyper planes in a high-dimensional space [12]. The accuracy of the SVM classifier improves significantly as the database size increases. Classification with SVM is accomplished by determining the best hyperplane that distinguishes between two classes of data points. The hyper plane that generates the greatest difference between two classes is the optimal hyperplane for SVM [14].

Consider a supervised SVM two class classification. i.e. binary classifier. Consider that the training data is represented by $\{p, q_i\}$, where $i = 1, 2, \dots, N$, and $q_i \in \{-1, +1\}$, where N is the number of training samples, $q_i = +1$ for class s_1 and $q_i = -1$ for class s_2 . It is possible to find minimum one hyper plane which is defined by vector v and bias b , which can separate the classes accurately as shown in Figure 4.

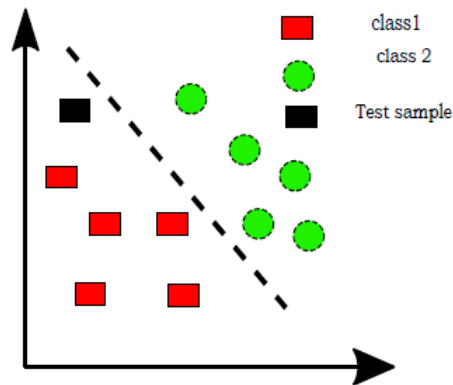


Figure 4. SVM Classification

4. Experimentation and Results

This study focuses on the identification of musical instruments. The most commonly utilized classical Indian instruments, such as the flute, harmonium, and sitar, are employed in this study. Initially, the sounds of these three instruments are gathered from standard recordings of them. These sounds have been preprocessed. An audio file of 3 seconds is obtained from the time frame in which the instrument is playing alone (monophonic) and converted to the .wav file format utilizing a file converter during preprocessing. The sampling frequency selected is 44.1 KHz. Audacity software is used to convert samples into 3 sec. samples. Some of the samples are prerecorded and some samples are recorded with instrument being actually played.

For this study we have considered 150 musical samples i.e. 50 samples each from flute, harmonium and sitar. We have divided this database into two i.e. training database and testing database. Training database consist 35 musical samples each from flute, harmonium and sitar i.e. total 105 samples. Testing database consist 15 musical samples each from flute, harmonium and sitar i.e. total 45 samples.

4.1 Spectral Feature of Monophonic Indian Musical Instruments

A total of 50 samples of each Musical Instrument (Flute, Harmonium, and Sitar) were studied. Feature vector for all 50 notes were extracted and its pattern were observed which appeared to be same. Table 1 and Table 2 shows Spectral values for Flute, and Harmonium as demonstrative case. Each table shows five samples of Instruments. The database and observations are given below.

Table 1. Database of Feature Vector of Flute

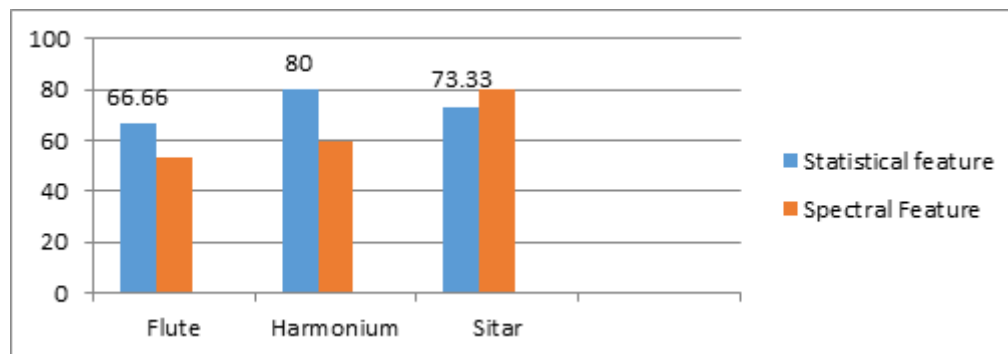
Sr. No.	ZCR	Periodicity	Amplitude	Brightness	BW	Roll off
1	3.69E-07	4487737	16.23962	-0.00709	0.016952	2885.649
2	2.67E-06	378670.2	100.672	0.00357	0.110547	18933.39
3	1.06E-06	341502.6	91.45122	0.003689	0.100359	17208.98
4	1.25E-07	11.03461	4.285959	-0.00086	0.004567	831.2392
5	2.29E-07	10.10459	4.85818	-0.00083	0.005136	949.5263

Table 2. Database of Feature Vector of Harmonium

Sr. No.	ZCR	Periodicity	Amplitude	Brightness	BW	Roll off
1	3.89E-07	0.075454	21.48429	-0.00122	0.026952	3538.32
2	4.68E-07	16.11275	18.44127	0.004069	0.024221	2907.583
3	4.83E-08	19.83793	9.497191	-0.00122	0.01191	1565.168
4	3.63E-08	6.140015	4.250988	0.00029	0.005474	682.2543
5	9.92E-08	130.347	6.887582	0.000221	0.008268	1185.594

4.2 Classifier Result

KNN is also a multi class classifier and accurately classifies the Indian Musical Instrument into three classes with an accuracy of 64.44% in Spectral domain & 72.92% in statistical domain. Figure 5 shows the results of classification using KNN classifier for Monophonic music of Indian Musical Instruments varying values of training-testing ratio. KNN seems to give consistent classification accuracy over varying range of training-testing ratio. We used Spectral & Statistical Features for recognition of Indian Musical Instruments.

**Figure 5. KNN Comparative Result**

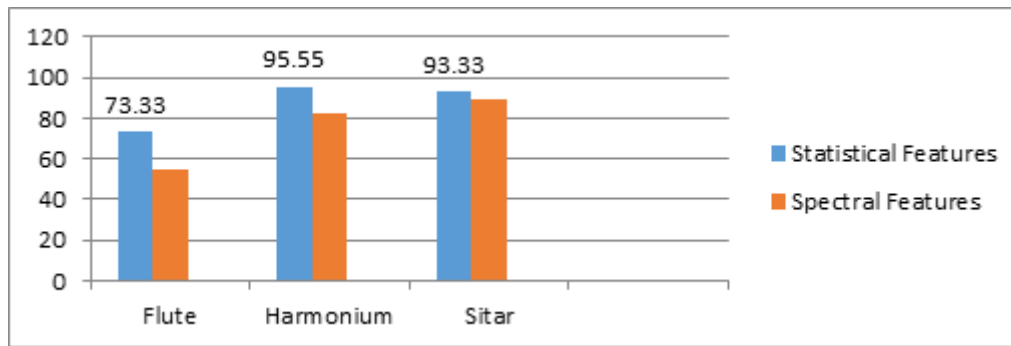


Figure 6. SVM Comparative Result

The above graphical representation describes the individual accuracy of each type of classifier. As depicted in Figure 6 and given in Table 3, the SVM binary classifier classifies the input sound sample of musical instrument into woodwind and string based class with a high accuracy of Harmonium 95.55%.

Table 3. Classification Accuracy for SVM

Monophonic		
Class	Statistical Parameter	Spectral Parameter
Harmonium and (Flute, Sitar)	95.55	82.22
Flute & Sitar	83.33	58.82

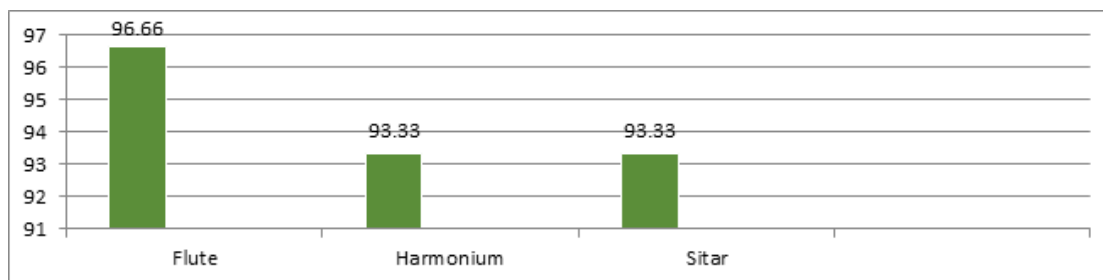


Figure 7. GMM Comparative Result

The above graphical representation in Figure 7, describes the individual accuracy of each type of classifier. The GMM classifier classifies the input sound sample of musical instrument into three classes with a high accuracy of Flute 96.55%.

4.3 Experimentation and Results (Polyphonic)

The objective of this stage of analysis was to identify number of instruments in input files and classify their type. This stage differ from previous stages of analysis as the number of sample files used for feature extraction were more in number and had multiple musical instruments embedded into it. Even though the accuracy obtained was a bit less as compared to previous stages of analysis, this final stage of experimentation helped in achieving exponential results in classification and identification of Indian musical instruments.

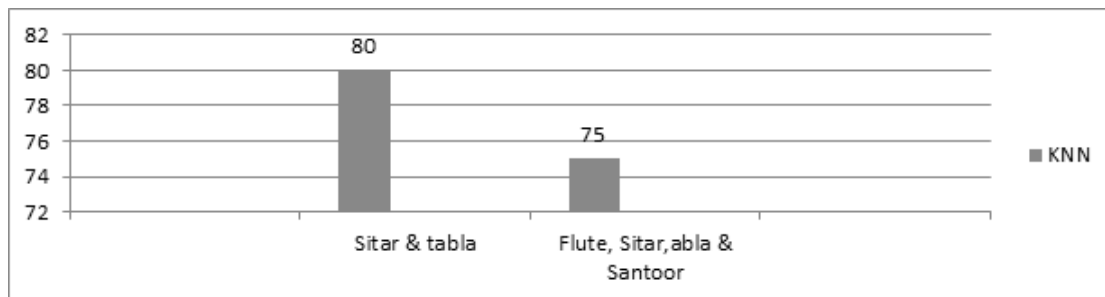


Figure 8. KNN Classification Accuracy for Polyphonic

KNN is classifier accurately classifies the Indian Musical Instrument into three classes with an accuracy of 80% in Spectral domain with two instruments and 75% in 75% with three instruments. Figure 8 shows the results of classification using KNN classifier for Polyphonic music of Indian Musical Instruments. While the graphical representation in Figure 9 describes the accuracy of SVM is substantially high for two instruments up to 93.33% and even for five instruments simultaneously it reaches to 88.88%. This is also represented in tabular form in the Table 4.

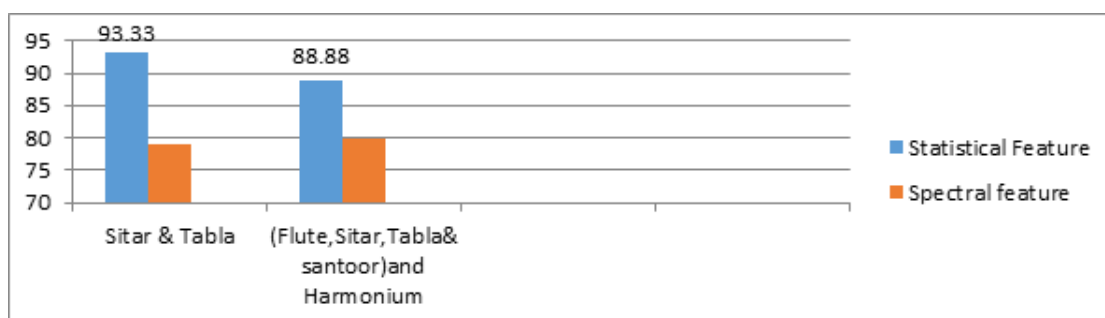
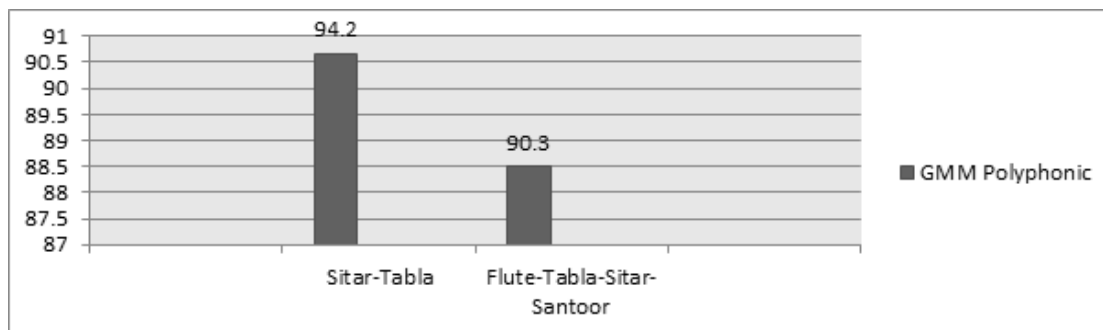


Figure 9. SVM Classification Accuracy for Polyphonic

Table 4. SVM Accuracy

Polyphonic		
Class	Statistical Parameter	Spectral Parameter
Sitar-Tabla	93.33	88.88
Sitar-Flute-Santoor-Tabla	88.89	80

The graphical representation in Figure 10 describes the accuracy of GMM. The GMM classifier classifies the input sound sample of two musical instrument with remarkable accuracy of 94.2 % and with four instruments it classifies with accuracy of 90.3%.

**Figure 10. GMM Classification Accuracy for Polyphonic**

5. Discussion and Conclusions

For monophonic musical samples KNN produces more accuracy in spectral domain than in statistical domain which reflects the fact that statistical domain parameters are more critical in instrument identification. Spectral parameters do contribute in deciding the quality of tone but as per as class of instrument is concern the statistical parameters play move vital role. This fact is also revealed by SVM which also demonstrates greater accuracy with statistical parameters. Being more effective SVM reaches accuracy of identification up to 95.5% specifically for Harmonium. For flute the behaviour produced by the air column generates a structure which resembles the memory generating system which is more complex than the harmonium. This is observed in the results.

GMM proves to be the best classifier for monophonic instrument identification. As GMM is mixture of multiple Gaussian distributions and the behaviour of the instrument note also resembles the Gaussian distribution when observed in the probabilistic domain enhances the result. Multiple Gaussian exhibit the most effective identification.

For Polyphonic SVM proves to be the best classifier with the accuracy of 94% for two instruments as the best performance of the SVM is reflected in binary classification. However

for more than two classes combined GMM supersedes SVM and gives the maximum accuracy of 88% for four classes combined. So up to two classes combined the SVM proves to be the better classifier but for more than two classes GMM proves to be the best classifier. Figure 10 graphically represents the comparison of classification accuracy for each Instruments. As can be concluded that, GMM outstands the other two classifiers and proves its efficiency by providing highest accuracy.

References

- [1] J. J. Burred, A. Robel and T. Sikora, "Dynamic Spectral Envelope Modeling for Timbre Analysis of Musical Instrument Sounds", IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 3, (2010) pp. 663-674.
- [2] Sang Hyun Park, "Musical Instrument Extraction through Timbre Classification", NVIDIA Corporation Santa Clara, CA 95050.
- [3] M. A. Bartsch and G. H. Wakefield, "Singing Voice Identification using Spectral Envelope Estimation", IEEE Transactions on Speech and Audio Processing, vol. 12, no. 2, (2004) pp. 100-109.
- [4] James W. Beauchamp. "The Sound of Music: Analysis, Synthesis, and Perception of Musical Sounds", pages 1-89, Springer, New York, (2007).
- [5] Daniel Mintz, "Toward Timbral Synthesis: A New Method for Synthesizing Sound based on Timbre Description Schemes", M.S. thesis, University of California, Santa Barbara, (2007).
- [6] J. F. Schouten, "The Perception of Timbre", Proceedings of the 6th International Congress on Acoustics, Tokyo, Japan, (1968) August 21-28.
- [7] T. H. Park, "Towards Automatic Musical Instrument Timbre Recognition", Ph.D. Thesis, Princeton University, (2004).
- [8] J. B. Allen, "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 25, no. 3, (1977) pp. 235-238.
- [9] Marques and P. J. Moreno, "A Study of Musical Instrument Classification using Gaussian Mixtures Models and Support Vector Machines", Compaq Cambridge Research Laboratory, vol. Tech. Report 99/4, (1999).
- [10] Swe Zin Kalayar Khine, Tin Lay New, Haizhou Li, "On Timbre Based Perceptual Feature For Singer Identification", Institute for Infocomm Research 21 Heng Mui Keng Terrace, Singapore.
- [11] Giovanni De Poli and Paolo Prandoni, "Sonological Models for Timbre Characterization", Journal of New Music Research, Vol. 26:2, (1997) pp. 170-197.
- [12] Christianini, N., and J. Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods", Cambridge University Press, Cambridge, UK, (2000).
- [13] B. Kostek, "Musical instrument classification and duet analysis employing music information retrieval techniques", Proceedings of the IEEE, vol. 92, no. 4, (2004) pp. 712-729.
- [14] Chih-Wei Hsu and Chih-Jen Lin, "A comparison of methods for multiclass support vector machines", IEEE Transactions on Neural Networks, vol. 13, no. 2, (2002) pp. 415-425.