OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

SONG RECOMMENDATION BASED ON VOCAL COMPETENCE

Ms. Samruddhi S. Doshi[1], Mr. Dinesh B. Hanchate[2]

Student, Computer Engineering, *VPKBIET, Baramati, India* [1]
Professor, Computer Engineering, *VPKBIET, Baramati, India* [2]

Abstract: Singing skills of each singer helps in choosing accurate songs for them. These skills are determined by using vocal competence. Matching a song to singer's vocal competence is the major and difficult task in song recommendation. There are other song recommendation systems, which works on recommending songs of listener's interest. These traditional approaches did not consider the singer's ability of singing which leads to poor singing performance by singer. This paper gives the solution using song recommendation framework. This framework uses the term singer profile, which constitutes singer's vocal competence in terms of features of voice like pitch, intensity and also the voice quality of the singer. The paper then gives technique to acquire singer profile and song profiles are stored in database. Then learning to- rank scheme is used to rank the recommended song profile for each singer profile.

Keywords: *Singing skills, Vocal Range Profile, Learning to-Rank.*

I INTRODUCTION

Singing skills of each singer helps in choosing accurate songs for them. These skills are determined by using vocal competence. Singers with different singing skills give better performance only if song is suitable to their skills. Many times, singers are upset due to their performance. The reason behind it is sometimes because of wrong preferred song by singer than the singing ability of singer. E.g. it is difficult task for a girl having soft voice to sing a song which requires strong voice. To achieve a good performance, songs are chosen strongly with singer's vocal competence.

Matching a song to one's vocal ability is the major task in song recommendation. Experienced singing teachers listens the voice of singer to find point in a voice. This point then used to choose suitable song matching with singer's voice. Typically, they choose tough songs for singers which differentiate singers from each other. They recommend songs which gives assurance of good singing performance.

This paper focuses on Song Recommendation Based on Vocal Competence. A good music recommendation system should be able to automatically detect preferences and generate playlists accordingly. Additionally, music recommendation system is used to help users to filter and discover songs according to their tastes. The main objective of this paper is to recommend challenging but feasible songs as par the vocal ability of the singer. Mainly the singer's digital voice is given as input to the system and analyzed for checking singer's vocal ability. According to competence search process is carried in the song database which recommends list of song. To the system singer's voice is treated as a query in the form of singer profile. For generation of singer profile vocal capability model called Vocal Range Profile (VRP) is used.

VRP of anyone in the (pitch, intensity) space, is a two dimensional bounded area. But the VRP is not sufficient to describe singing competence because it overlooks the voice quality. So system first generates VRP and then calculates voice quality as a numerical function of pitch-intensity space. Singer profile then consists of VRP of singer and its respective voice quality. The database of the system consists of number of song profiles. Similar to singer profile, song profile represents pitch and intensities of music notes. So for recommending songs, system needed to work for matching singer profile with song profiles. System can use approach of ranking based on competence. This scheme is used for ranking the songs in the dataset for singer. This recommendation is different because this does not recommend songs matching with listener's interest. The main advantages provided by this paper:

1. To enable singers sing songs according to their vocal ability.

2. To help singing teachers to recommend and rate the song.

The research purpose of the song recommendation system is slightly differentiating from most existing music fetching systems. Historical systems only focus on extracting song with listener's interests. The methods previously used for music retrieval uses content based music retrieval [2], [3], [4] techniques. Other techniques based on collaborative methods recommend a song in a group of users having the same interest.

Terminologies: This section defines and introduces relevant key terms as used in the article. These terminologies help in understanding recommendation procedure. A vocal range profile (VRP) is also referred as a phonetogram. It is a two dimensional map where each point represents the vocal ability of a human. This map depicts all possible (pitch, intensity)- pairs that any human can produce. Arranging a VRP map on the pitch axis results in the range of pitches that any person can at all produce, is referred as pitch range. Generally, the VRP specifies the voicing ability of anyone by depicting the highest and smallest vocal intensity at each pitch value over the complete pitch range.

II REVIEW OF LITERATURE

Traditional approaches of song recommendation investigate in field of content basis song recommender systems or used collaborative techniques for recommending particular song. These techniques find listener's favorite song concerning music content correspondence such as moods and rhythms. Collaborative techniques recommend a song in a group of users having the same interest. Content based music retrieval methods are given below.

In [2], K. Hoashi, K. Matsumoto, and N. Inoue given a music retrieval method which used to retrieve songs as per the users singing priorities. This permits the users to find new songs according to their interest and also they are expected to like. The given implementation of relevance feedback method is for improvement of the performance of music information extraction method. The burden on users of inputting learning data to the system has been reduced with use of method which generates user profiles based on grouping preferences. Also they did refinement process for such profiles using relevance feedback. Another technique of content based music retrieval, given in [3]. They have proposed the system to construct music descriptors. That system supported the efficient content-based music extraction and categorization. This method has given integration of multiple music feature vectors into a single low dimensioned vector, based on characteristics as timbral texture, pitch, and rhythm structure. The work of generating music descriptor carried in two stages- a. Dimension reduction using Principal Component Analysis (PCA), b. Non-linear neural network used for generating music descriptors.

The next method of content based music retrieval is Scalable Content-Based Music Retrieval [4]- This method uses Chord Progression Histogram and Tree-Structure LSH for Music fetching. As multimedia content increasing day by day over the Internet, music data extraction has become difficult task. This method uses the melody similarity for quick and reliable retrieval of relevant songs. The system works in two phases: firstly, representation of audio tracks done correctly using music semantics. Then chord progressions has been identified from audio signals and for improvement of accuracy of identification, method has used multi-probing. Afterwards, as a mid-level feature, concise chord progression histogram (CPH) is computed from each audio track, which is used to describe audio content. As a second step it organized audio tracks as par their CPHs for this it used one locality sensitive hash table
with a tree-structure.

The techniques of content based method only focused on the song recommendation as par the listener's interest. Consideration of the singer's ability is main aspect in song recommendation which has not covered in earlier studies. Technique given in [5], proposed a song suggesting framework for making song recommendation in public community. Despite of suggesting tunes which audience like to hear, this system used to suggest feasible songs that people can sing efficiently. System has proposed with the aim of finding the song difficulty orderings among the song performance ratings of each user. The system transformed the difficulty orderings into a difficulty graph. Iterative inference algorithm was proposed to suggest songs relevant to the difficulty graph. To evaluate the singer's performance, technique proposed by W.-H. Tsai and H.-C. Lee [6], with objective of developing an automatic singing evaluation system for Karaoke performances. This technique has used vocal features like volume, pitch and rhythm to assess a singer's performance and improved the singing evaluation capabilities on Karaoke machines. The performance of singer calculated concerning technical accuracy and system then assigns a rating score. It produces the results as close to the human rating.

The main task in music retrieval is feature extraction. One of the approach of feature extraction is given in[7]. The method is given to extract features for purpose of key detection. In music a key a set of musical notes that are primarily used for constructing a bit of music. The key gives dominant statistics regarding the musical content of a music piece, e.g. like harmonic and melodic context. The system has used consonance filtering for extraction of features that used for key detection in music.

 ISSN (Online) 2456-3293

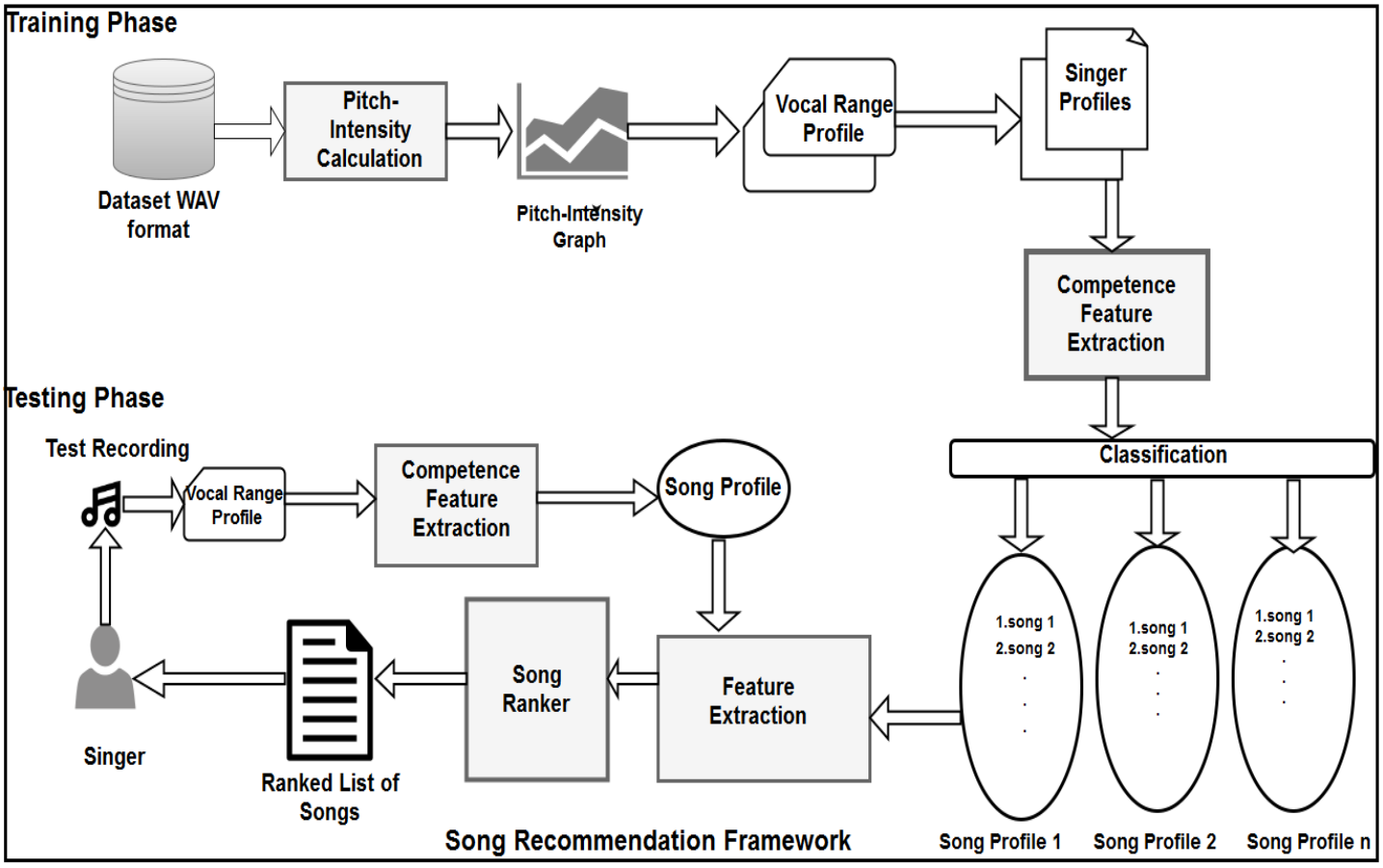

Figure 1 : Song Recommendation Framework

E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri [8] given techniques is about music transcription. This is the operation of transforming an audio into some form of musical notation. The vital problem considered in AMT system was the diagnosis of various concurrent pitches. For selection of feature which helps in evaluating singer's performance, techniques are required. One of the techniques is given in [9]. This method proposes a feature selection (FS) algorithm to filter the low efficiency features for fast speech emotion recognition. They have proposed an algorithm to select features by considering discriminative ability as well as time consumed and redundancy of each feature. Technique given in [9] reduces the dimensionality of qualified features and at the same time refine its discriminative ability for a more systematic and effective emotion recognition system. To retrieve songs matching with singer's vocal competence from database ranking is done. Technique described in [10], gives the method of ranking. In this technique of ranking, on inputting a query, the ranking function allot a rating to each document. It does the positioning of documents in descending order of the rating. The ranking order represents the similarity of documents with respect to the query. This can also used for ranking of songs in music retrieval.

III SYSTEM OVERVIEW

The system takes the data for recommending songs to particular singer as their vocal samples i.e voice samples of singers. These samples are treated as singer profiles. From the singer profiles system extracts the features and determines voice quality of singer. Using that voice quality of singer profile, system will check the dataset for matching song profiles. When system found the matching profiles it will return that song profiles according to their ranking in one list called ranking list. This list gives the song which are suitable to particular song profile. In this way, system carries the work of song recommendation. Figure 1 shows the architecture of recommendation system.

A. Singer Profile

A singer profile [1], contains two components:

1. VRP of the singer and
2. A voice quality function defined over the VRP area.

B. Singer Profile Modeling

A singer profile is defined as a tuple of $< VRP, \psi >$

Where, VRP is the VRP of the singer and ψ is singer's respective voice quality function.

C. Competence Feature Extraction

In song recommendation, voice quality of individual singer plays a significant role. For evaluation of Voice quality, pitch-intensity differences of different singer profiles are considered, which are calculated using Euclidian distance. If difference is less quality is a good and quality decrease as difference increases. This evaluation done using feature given in Table 1.

Table 1 Features

Feature Category	Feature Name
Pitch Features	medianPitch, meanPitch, sdPitch, min-Pitch, maxPitch, nPulses, meanPeriod, sdperiod
Frequency Perturbations	Jitter_loc_abs[11], jitter_loc, jitter_rap[11], jitter_apq5[11]
Amplitude Perturbations	Shimmer_loc[11],shimmer_locdB[11], shimmer_apq5[11], shimmer_apq3[11], shimmer_apq11[11]
Spectrum Features	Mean_nhr[12], mean_hnr[13],singing power ratio[14], tilt[12], ltasS_lope[15]

D. Classification

After generating the singer profiles, system need to classify them into accurate singer profiles. For that purpose, firstly similarity between singer profiles are calculated using formula:

$$sim(song_0, song_i) = \frac{\sum(song_{0(v_j)} * song_{i(v_j)})}{\sqrt{\sum song_{0(v_j)}} * \sqrt{\sum song_{i(v_j)}}} \quad (1)$$

Where,

- $song_0$ is the query song and $song_i$ are the songs in available class.
- v_j is the jth feature vector of $song_0$ and $song_i$.

If there are more than one song in the class, system will calculate the average similarity of the class. This procedure is given in algorithm(1).If this calculated similarity is greater than threshold then we allocate flag to matched class with the value of similarity, and matched class will appear as nearest class of query song. If class with higher similarity found, flag and nearest will change otherwise song will added to nearest class. This procedure is given in algorithm (2).

E. Recommendation and Aggregation of results

For recommending songs, system uses machine learning. It takes three different methods of machine learning like Naïve Bayes, Neural Network and ID3 (i. e. Decision Tree). For this purpose, system creates training instance of every class generated during classification of dataset. This instance stores VRPs and their occurrences in particular songs in the format of ARFF (Attribute Relationship File Format). This instance also stores classes of every song in the dataset. This training instance used afterwards for comparison with testing instance of testing song, which is also stored in ARFF file format. The individual results are also aggregated for high accuracy.

Input: List of Songs having TF, Threshold Thr = 0.5;
Output: List of classes where, $Class_c$ =List of songs.

$Class_0 \leftarrow Song_0$
for each remaining song $Song_r$ do
nearest = =-1;
flag = 0;
for each class Classc do
sim = *CalculateSimilarity*($Class_c$, $Song_r$);
if ($sim \geq thrANDsim \geq flag$) **then**
nearest = c;
flag = sim;
end
if (nearest==-1) then
Create new class $Class_n$;
$Class_n = Class_n \cup Song_r$;
else
$Class_{nearest} = Class_{nearest} \cup Song_r$;
end
return class;

ALGORITHM 1: PROFILE CLASSIFICATION

avgsim = 0;
for each $Song_i \in Class_c$ **do**
vector1 = TF_{Song_i} ;
vector2 = TF_{Song_r} ;

$$sim(song_i, song_r) = \frac{\sum(song_{i(v_j)} * song_{r(v_j)})}{\sqrt{\sum song_{i(v_j)}} * \sqrt{\sum song_{r(v_j)}}};$$

avgsim+ = sim;

$$avgsim = \frac{avgsim}{Class_c.length};$$

end

return avgsim;

ALGORITHM 2: SIMILARITY CALCULATION

It is challenging task to recommend the songs according to singer's vocal ability. There are methods which recommend song which are matching with listener's interest. For recommending songs according to singer's vocal ability system requires singer's vocal ability which is extracted from song using VRP and Voice Quality. The song recommendation system based on singer's vocal competence is mainly based on how efficiently singer's vocal ability is learned by the system. For learning purpose system took approach of List net which requires some features for learning. So learning method is divided into:

1. ListWise Approach[1]:

Within ListWise approach, Song Ranking system refers profiles of singer as a input query, and the song profiles are referred as documents. This approach aims at learning a function of ranking that taking feature vector V defined over every <singerprofile, songprofile> pair as input. Then this approach is used to produce ranking values of the songs.

2. Feature Extraction:

Given $< \sin gerprofile, songprofile >$, for any area Ar in the singer profile, assume that $\{term_1, term_2, ..., term_n\}$ are song terms occurring in Ar and their $term_{freq}$ in are denoted by $\{tf_1, tf_2, ..., tf_n\}$. Using this we can calculate features on area Ar as follows:

1. Total TF: This feature is calculated as $\sum_{i=1}^{n} tf_i$.
2. Total TF-IDF: Total TF-IDF of area Ar is calculated as $\sum_{i=1}^{n} tfidf_i$.

F. Competence Based Song Ranking

IV THE DATASET

As a dataset we took number of songs from website Indiamp3. com and remove music to extract core voice from song. We have collected 50 songs (25 for male, 25 for female) as the training dataset. Then convert them into wav format using Wave Pad Sound Editor. Then calculate the pitch and intensities of every singer's from their test song samples.

V RESULT ANALYSIS

A. Data Discussion

Table 2 Data Analysis

Sr. No.	# Songs	# VRP Profiles
1.	9	2000
2.	16	2958
3.	12	2246
4.	14	2500
5.	17	3134

As shown in Table 2, firstly we take a set of 9 songs, with high variation then extract features from song we get 2000 VRP profiles. Then we increase the number of songs for processing on system. At second step we increase 7 song without increasing variation in song types. In the third step, we decrease 4 songs from previous set of songs, this shows effect on VRP profile we got less no. of VRP profiles. Then again we start increasing the number of songs and decrease the variation to check the efficiency of the system.

Table 2 Result Table

Sr.No	# Songs	# Classes	# Relevant Songs	#Recommendations	# Correctly Recommended	Precision	Recall
1.	9	4	6	4	3	0.67	0.75
2.	16	5	7	6	5	0.83	0.75
3.	12	4	6	5	4	0.8	0.83
4.	14	5	7	6	5	0.83	0.85
5.	17	3	16	14	12	0.86	0.88

ISSN (Online) 2456-3293

B Result Table

For each song set, system generates number of singer profiles, which depending on features, generate classes and recommend list of songs. E.g., for set of 9 songs having large variation in song types as shown in Table III and for a given query there are 6 relevant songs in dataset and system could generate 4 recommendation out of them 3 are correctly recommended. At the second stage, we increase the number of songs as well as do not increase the variation in song types. At this step system gives good performance. Then in third stage, we keep the same variation and decrease the number of input songs. In the last two cases, we increase the number of input songs and decrease the variation. At both the cases system is having better performance. In both first two cases, there is low performance of system. On the basis of this we can do result analysis.

To calculate the performance measures, we use the above formula. To calculate precision we can decide correctly recommended song with actual number of recommendations. Similarly to calculate recall, we can divide number of recommendations with number of relevant songs in the dataset. Precision and recall is calculated using following formula:

$$\mathrm{Pr}ecision = \frac{\#Correctly\ \mathrm{Re}commendedSongs}{\#\mathrm{Re}commendations} \quad (2)$$

$$\mathrm{Re}call = \frac{\#\mathrm{Re}commendations}{\#\mathrm{Re}levantSongs} \quad (3)$$

C Result Graph

By using the precision and recall measures, we do result analysis. Figure 2 shows the result graph, by viewing graph we analyse the system. By analysing the graph we can say that as there is high variation in song type affect the performance of the system. The last case shows that even there is large load on the system, due to less variation, system perform greatly.

A. Training Phase

1) Firstly, dataset of songs present in WAV format are browsed as shown in figure 3.

2) Then pitch and intensities and VRPs of each song is calculated. Then Classification of every song is done by finding similarity of every song with every other song in dataset. This is shown in figure 4.

3) For training we firstly create training instance of every class and then use three machine learning algorithms like neural network, naive bayes, decision tree as shown in figure 5.

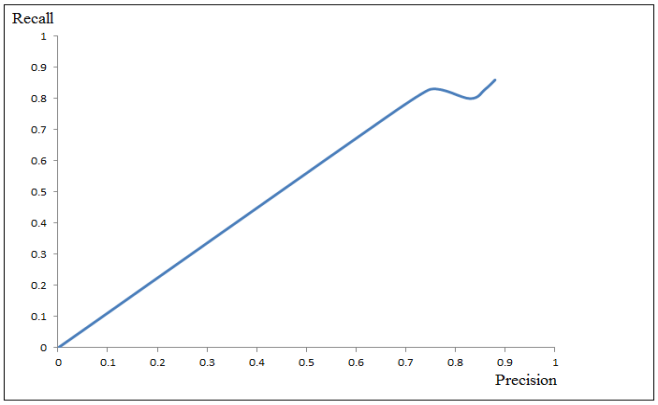

Figure 2 Result Graph (Precison-Recall)

B. Testing Phase

1) In testing phase firstly, browse the query song present in WAV format are browsed as shown in figure 6.

2) Then pitch and intensities, VRPs of query song is calculated. And finally after predicting the class, it will recommend the songs with ranking to the user as shown in figure 7. Here, results of naive bayes are only shown.

3) Aggregated results are also displayed as shown in figure 8.

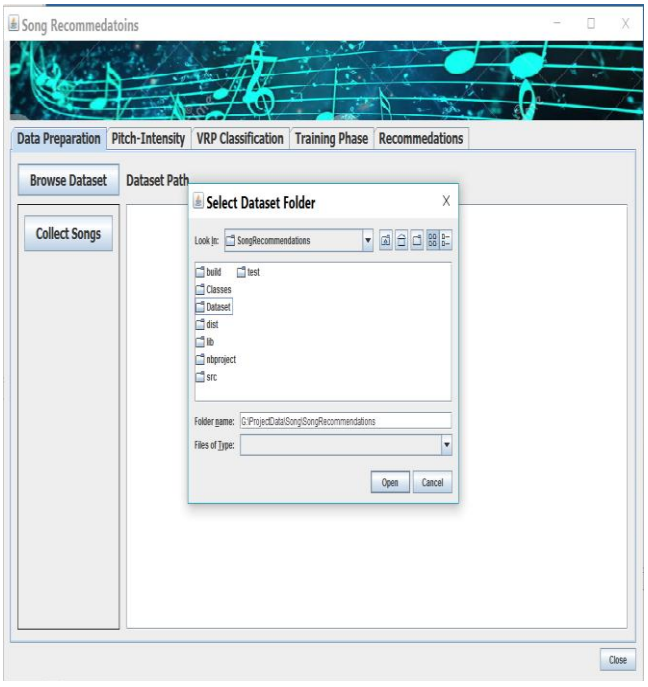

Figure 3 Browse Dataset of songs in WAV format, to calculate pitch-intensities. These pitch-intensities later used for plotting VRP graph.

ISSN (Online) 2456-3293

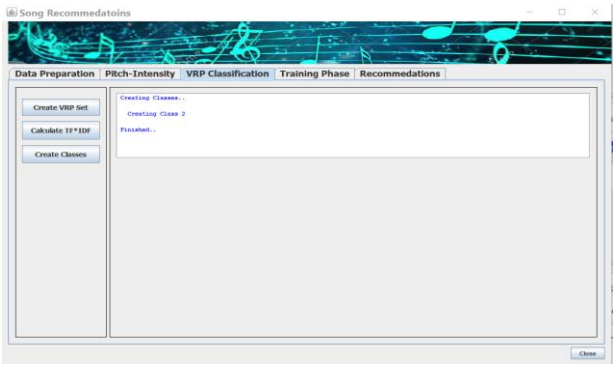

Figure 4 Classification of VRP profiles according to their respective voice quality.

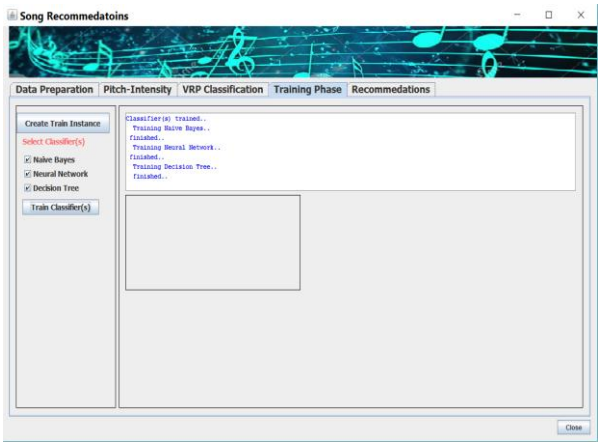

Figure 5 After classification, every song of every class is trained and its instance is stored in ARFF format.

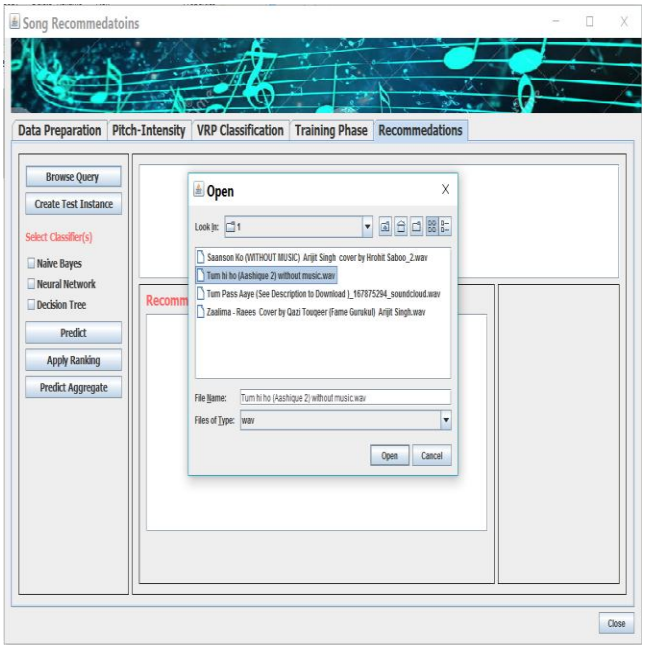

Figure 6 Query song is browsed for finding its respective matching profile.

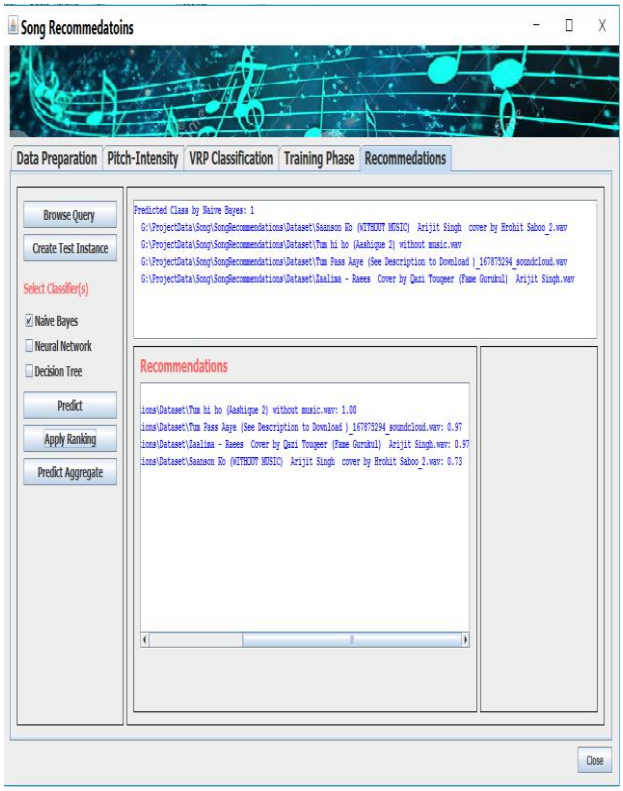

Figure 7 VRP graph is plotted of test song as per the pitch-intensity values. Then it is matched with most appropriate singer profile. And then it will be returned to user with ranking.

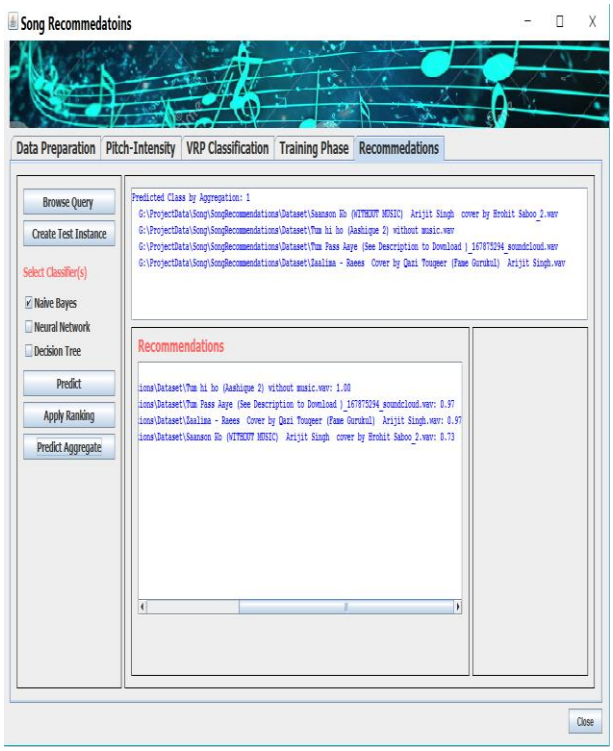

Figure 8 Aggregate Results of three classifiers that are calculated for more accuracy.

VI CONCLUSION

In the Song Recommendation system, vocal competence of singer is modeled as singer profile. Singer profile takes (pitch, intensity) pair and also the quality of voice into the consideration. Then using voice quality we divide the number of singer profiles in to different classes (i. e. Song Profiles). One of the song profile is recommended to user on given query. In section result analysis, system is checked for different inputs with load and increased variations. As Table III represent the result analysis and from the result graph it is clear that system give accurate results when there are less variations.

VII FUTURE WORK

In the future work system takes the singers choice of song into the consideration. In this system, singer may be unhappy with recommendation of system because, song recommended to singer is considering the capability only but not considering the choice of singer. So this system in future will also consider the singer's choice. And also the system will work for taking listener's interest. Currently system works only for recommending songs according to singer's quality, but system do better recommendation by considering singer's choice and listener's interest.

ACKNOWLEDGMENT

This is to acknowledge and thank all the individuals who played defining role in shaping this work. I avail this opportunity to express my deep sense of gratitude and whole hearted thanks to Dr. Dinesh B. Hanchate for giving his valuable guidance, inspiration and encouragement to complete this task.

REFERENCES

[1] Kuang Mao, Lidan Shou, Ju Fan, Gang Chen, and ohan S. Kankanhalli, "Competence-Based Song ecommendation: Matching Songs to Ones Singing Skill,"IEEE Transaction on Multimedia, vol. 17, no. 3, mar. 2015.

[2] K. Hoashi, K. Matsumoto, and N. Inoue, "Personalization of user profiles for content- based music retrieval based on relevance feedback," in Proc. ACM Multimedia, 2003.

[3] J. Shen, J. Shepherd, and A. H. H. Ngu, Towards effective content-based music retrieval with multiple acoustic feature combination, IEEE Transaction on Multimedia, vol. 8, no. 6, pp. 11791189, Dec. 2006.

[4] Y. Yu, R. Zimmermann, Y. Wang, and V. Oria, "Scalable content-based music retrieval using chord progression histogram, and tree-structure LSH," IEEE Transaction on Multimedia, vol. 15, no. 8, pp. 19691981, Dec. 2013.

[5] K. Mao, J. Fan, L. Shou, G. Chen, and M. S. Kankanhalli, "Song recommendation for social singing community," in ACM Multimedia, 2014.

[6] W.-H. Tsai and H.-C. Lee, "Automatic evaluation of karaoke singing based on pitch, volume, and rhythm features," IEEE Transaction on Audio, Speech, Lang. Process., vol. 20, no. 4, pp. 12331243, May 2012.

[7] Y. Zhu and M. S. Kankanhalli, "Precise pitch profile feature extraction from musical audio for key detection," IEEE Transaction on Multimedia, vol. 8, no. 3, pp. 575584, Jun. 2006.

[8] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, "Automatic music transcription: Breaking the glass ceiling object recognition and segmentation," in Proc. ISMIR, 2012, pp. 379384.

[9] L. Zhang, M. Song, N. Li, J. Bu, and C. Chen, "Feature selection for fast speech emotion recognition," in Proc. ACM Multimedia, 2009, pp. 753756.

[10] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li, "Learning to rank: From pairwise approach to listwise approach," in Proc. ICML, 2007.

[11] P. Boersma and D. Weenink, Univ. of Amsterdam. Amsterdam, The Netherlands, Voice, Oct. 2006 [Online].

[12] D. Deliyski, Acoustic model, and evaluation of pathological voice production, in Proc. Eurospeech, 1993, pp. 19691972.

[13] E. Yumoto, W. Gould, and T. Baer, Harmonics-to-noise ratio as an index of the degree of hoarseness, J. Acoust. Soc. Amer., vol. 71, no. 6, pp. 15441550, 1982.

[14] C. Watts, K. Barnes-Burroughs, J. Estis, and D. Blanton, The singing power ratio as an objective measure of singing voice quality in untrained talented, and nontalented singers, J. Voice, vol. 20, no. 1, pp. 8288, 2006.

[15] G. Peeters, IRCAM, Paris, France, A large set of audio features for sound description, Tech. Rep., 2004.