EURASIP Journal on Audio, Speech, and Music Processing

RESEARCH **Open Access**

Robust singer identification of Indian playback singers

Deepali Y. Loni[1*] and Shaila Subbaraman[2]

Abstract

Singing voice analysis has been a topic of research to assist several applications in the domain of music information retrieval system. One such major area is singer identification (SID). There has been enormous increase in production of movies and songs in Bollywood industry over the last 50 decades. Surveying this extensive dataset of singers, the paper presents singer identification system for Indian playback singers. Four acoustic features namely—formants, harmonic spectral envelope, vibrato, and timbre—that uniquely describe the singer are extracted from the singing voice segments. Using the combination of these multiple acoustic features, we address the major challenges in SID like the variations in singer's voice, testing of multilingual songs, and the album effect. Systematic evaluation shows the SID is robust against the variations in singer's singing style and structure of songs and is effective in identifying the cover songs and singers. The results are investigated on in-house cappella database consisting of 26 singers and 550 songs. By performing dimension reduction of the feature vector and using Support Vector Machine classifier, we achieved an accuracy of 86% using fourfold cross validation process. In addition, performance comparison of the proposed work with other existing approaches reveals the superiority in terms of volume of dataset and song duration.

Keywords: Vibrato, Formants, Pitch, Singing voice, Singer identification, Spectral envelope, Timbre

1 Introduction

A singing voice is the basic element of a song. Every singer's voice possesses certain specific acoustic features which mark their individuality. These unique characteristics of the singers can be utilized for various music information retrieval (MIR) applications like singer identification, singer verification, identification of trained and untrained singers, separation of singing voice from accompanied music, singing voice detection, signal enhancement, and many more [1].

The source of the singing voice is the vibrating vocal folds that produce the air pulses in the vocal tract which generates the sound [2]. The vibrating vocal folds determine the fundamental frequency (F0). The singer's voice is therefore characterized by the physical structure of the vocal system like the shape, size and muscularity of vocal tract; the jaw movement (opening); the body and tip of the tongue; the larynx position; and the expansion of pharynx. Studies reveal that singers consciously or unconsciously change the acoustic parameters to match either the style of singing or the accompanying music, to ornament his/her tone or to affect singing voice perception [3, 4]. Also, very often, singing is under the strong influence of accompanied music which interferes with the singer characteristics [5]. Moreover numerous physiological changes occur in the singer's voice with the aging process, and these changes have a significant effect on the acoustical features defining the singing voice [6]. These wide governing factors make singer identification (SID) a challenging task.

Each acoustic feature of the singing voice describes the characteristic of a particular vocal organ. Shen et al. [7] found that human perceptual system interprets and processes a music signal using various kinds of acoustic features. Therefore, a single type of acoustic feature does not provide effective information to represent the singing voice broadly. In this paper, we propose a new approach for singer identification that exploits the combination of four major acoustic features of singing voice, namely, formants, harmonic spectral envelope, vibrato, and timbre. From the pitch of singing voice, we

* Correspondence: deepaliloni@rediffmail.com
[1] D.K.T.E's Textile & Engineering Institute, Ichalkaranji, India
Full list of author information is available at the end of the article

characterized the vibrato (undulation of F0) and harmonic spectral envelope features. Formants are governed by the shape of vocal tract whereas timbre describes the spectral characteristics which depend on shape and size of the vocal cavities, i.e., spectral content of singer's voice. All these acoustic features closely describe the singing voice production, i.e., the vocal system comprising the respiratory parts, the vocal folds, and the vocal tract. Consideration of these features together strongly account to precise description of singing voice.

The songs of Bollywood are popular across the globe and are the most often searched items on the web from India [8]. Bollywood film songs have been described as eclectic both in instrumentation and style [9]. But there has been little research contribution on Indian film songs and singers. As a first step in our research work, we created the database of cappella singing voice of popular singers from Indian film industry. The proposed approach of SID can be used to cluster songs of similar voices of singers in a music collection, or search the singer of a query song. We have tested our SID approach using a complete non-homogenous dataset which includes songs of wide time span for each singer, thus capturing maximum variations in the singer's voice. Experimental results show that our proposed system is robust and independent of song genre, language, and style of singing.

The rest of this paper is organized as follows. Section 2 describes the related work. Section 3 discusses the singer identification system, the acoustic features, and their extraction process. The cappella database of Indian playback singers used in the experimentation is described in Section 4. In the results section of the paper, we discuss effect of song duration on SID and compare the performance of the classifiers for different principal component analysis (PCA) dimension. Experimental results demonstrate that our SID system outperforms in terms of volume of data, song duration, and SID accuracy when compared to other state-of-the-art approaches using Indian song database.

2 Related work

Singer identification task initially was attempted by applying the existing speech processing techniques. Whitman et al. [10] attempted SID for a database of roughly 210 songs covering 21 artists using frequency content features, which reported accuracies of approximately 50%. Singer identification developed by Tong Zhang [11] used LPC Cepstral coefficients of each audio frame; however, it was observed that the accuracy rate is not consistent among the singers. Mel frequency cepstral coefficients (MFCC) are a vocal-related acoustic feature that is the most commonly used speech recognition tool which has been explored extensively for SID [12–17]. The problem observed with MFCC is that it fails to retain the formant information of high pitched singing voice as the harmonic components become sparse.

Kim and Whitman [18] proposed formant frequencies and magnitudes analyzed via warped linear prediction as key features to distinguish singers' voices from one another. The system of Shen et al. [7] considered spectral centroid, spectral flux, spectral roll, zero crossing rate, energy, and MFCC to form the timbre vector along with rhythm and pitch features for the task of efficient music retrieval, but the descriptors used for SID task are not well defined. Other than SID, timbre has been explored for other music-related research activities like vocal detection [19], distinguish different types of musical instruments or voices [20] etc. The investigations in [5, 21] proved the effectiveness of vibrato feature in describing the singing-voice characteristics. The most commonly observed vibrato extraction process includes formulating the F0 contour, where its spectral peak corresponds to vibrato rate while the vibrato extent corresponds to F0 excursions [22–25]. The methods differ in the approach of identifying and separation of vibrato segments; some perform manual annotation while others do it automatically. Bonjyotsna and Bhuyan [26] observed that the vibrato feature of Indian singers has not yet been analyzed much. Mesaros and Moldovan [27] proposed SID using energy coefficients as features determined using a Mel-scaling of the FFT-bins and other method derived from a fractional B-spline-based decomposition.

Jialie et al. [28] developed a hybrid model for singer identification using multiple low-level features extracted from both the vocal and non-vocal music segments. The approach considered that the styles of song performed by a singer could be relatively steady during a certain period. However this assumption may not be true to all singers, especially for Indian songs. Tsai and Lee [12] proposed an entirely different approach for SID. The approach investigated the singer identification by combining the features of singer's voices along with singer's spoken data. Experimental results indicated that the system trained using only speech utterances performed rather poorly in identifying singers, compared to that of the system trained using cappella singing recordings.

In regard to Indian music, the two broad categories of Indian songs are the Hindustani (North Indian style) and the Carnatic (South Indian style). Carnatic singer identification was proposed for six singers by Sridhar and Geetha [29] for efficient music retrieval of Carnatic-based songs. Saurabh et al. [30] identified ten singers from North Indian classical music using timbre descriptors for noise-free studio recordings accompanied with continuous background music of tanpura and violin/harmonium/flute, etc., and achieved an accuracy of 58.33%. A technique of identifying singers using video songs from the Internet and CD for Indian playback singers using cepstral coefficients is presented in [31, 32], however tested for a lower corpus. It is quite obvious, as the number of music items (songs and singers) increase, the performance of the system degrades due to noise and presence of more similar songs in the database [28, 33]. The research approach by Patil et al. [34] used cepstral mean subtracted MFCC for identifying 20 Indian playback singers, but operated on a large song duration of 60–166 s.

From the view point of extending the dimension of singer identity, we propose to investigate an approach that combines multiple acoustic features of singing voice. Systematic evaluation and the approach adopted for each acoustic feature extraction is presented in this paper. Formant frequencies are investigated using the combination of wavelet–LPC technique. We constructed the harmonic spectral envelope feature from the pitch of singing voice and found it to be unique for every singer. Identifying the vibrato sections from the singing segment, we extracted the vibrato parameters (rate and extent). We analyzed different spectral timbre features and selected spectral roll-off to characterize the timbre feature.

3 Singer identification system—the methodology

In this section, we present the singer identification system. The proposed experimental work extracts four acoustic features to identify the singer from the given segment of singing voice. The system operation is divided in two phases: training and testing. The training phase is the most crucial phase in any identification based system. In the training phase, we compute the acoustic features and analyzed them to find distinct characteristic range of each feature for every singer. For feature extraction, we used cappella singing segments of 3-s duration. Each analyzing singing voice segment is divided into frames of 25 ms, with 10 ms of overlapping in the preprocessing stage. The frames are hamming windowed to reduce the signal discontinuities. The features extracted from all the training samples of each singer are then assembled to form the feature vector to train the classifier. The entire structure of the SID system is shown in Fig. 1, and the information of each component in the framework is discussed in detail.

3.1 Formants of singing voice

The formants are the natural resonances of the vocal tract. The shape of vocal tract determines the location of formants, reflected as the peaks in the spectral envelope of the singing segment. The frequencies corresponding to these peaks are called the formant frequencies which characterize the singing voice. To compute formants, we used a combination of wavelet transform with linear predictive coding (LPC) technique. This combination overcomes the drawback of standard LPC approach. LPC weighs all frequencies equally on a linear scale, while the human perception characteristics are very close to logarithmic [35]. Standard LPC can miss closely spaced resonant peaks. The multi-resolution capability of

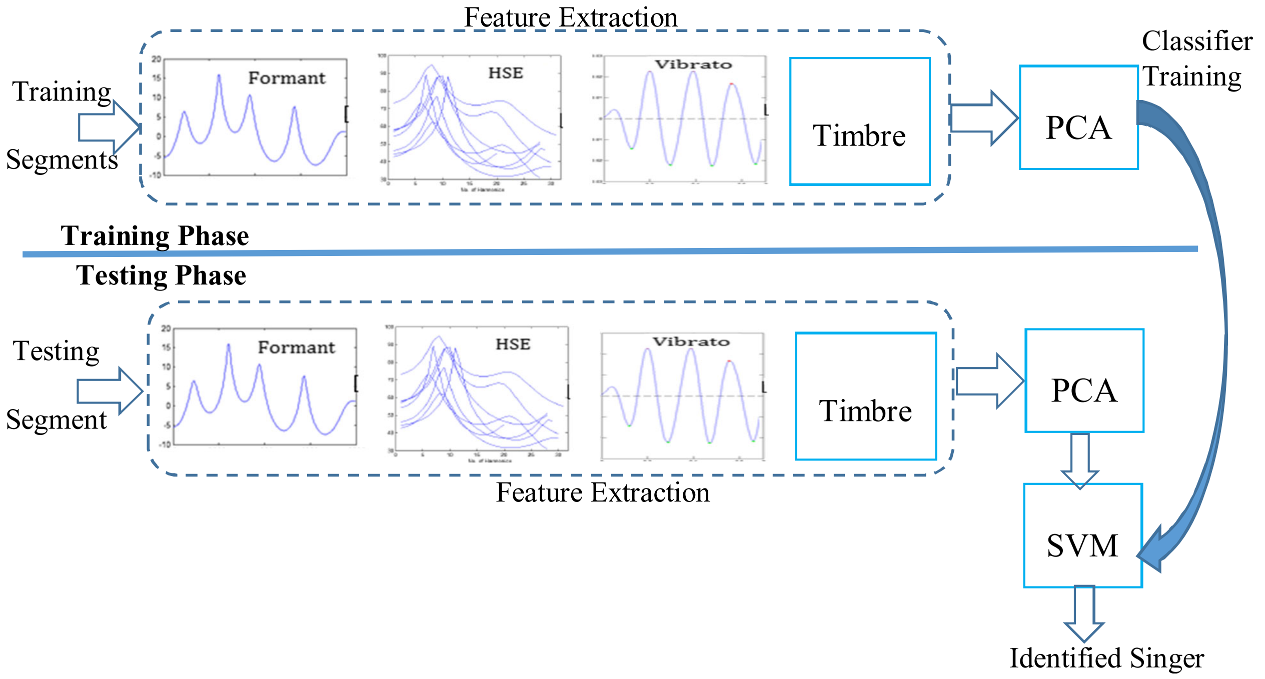


Fig. 1 Singer identification system

wavelet transform provides high frequency resolution to precisely capture formants from each subband according to their perceptual importance. The spectral peaks corresponding to the maximum amplitude are extracted as formants from each subband. The formant estimation process is presented in Algorithm 1.

```
Algorithm 1: Formant estimation

Step1 – Set threshold for formant estimation.
Step2 – Perform framing of input singing segment of 3 sec duration
        (Frame duration – 25msec, Frame overlap – 10msec, Sampling
        Frequency ($F_s$) – 44.1KHz).
Step3 – Perform wavelet decomposition of each frame (Wavelet – Symlet,
        No. of decomposition levels – 4)
Step4– Compute LPC coefficients (rts) from each subband.
        (LPC order =12)
Step5– Compute formants from LPC coefficients applying set threshold.
        $Formants = \frac{F_s}{2\pi} tan^{-1} \frac{Im(rts)}{Re(rts)}$
        $|rts| > threshold$
Step6 – Repeat the Steps 3, 4 & 5 for all the frames.
Step7 –.Select the formants from the subbands with higher peak strength.
```

There are generally five formants relevant to singing, wherein, contribution to overall projection of sound are believed to be made by formants higher than the third [36]. To analyze this, we computed the formants for different conditions on the roots of LPC coefficients and observed the effect on the formants. Figure 2 shows the result of formant analysis for a particular singer. We observed that for roots closer to unit circle (i.e., higher thresholds), LPC discarded lower formants while retaining only higher order formants. It indicated that higher formants more prominently contribute to define the overall projection of singing voice. For singer identification, we considered the higher formants which are retained by the system for the threshold value of 0.98. We considered 16-dimensional vector to represent formant information, containing formant frequencies F3, F4, F5, and F6, their mean and variance along with their peak amplitudes.

3.2 Vibrato of singing voice

One of the widely used quality measuring benchmark of singing is vibrato. It is considered as a good singing style acquired by the singers after many years of extensive vocal training [21]. Vibrato of singing voice is an oscillatory effect produced due to periodic variation in pitch of a sustained note. It is the glottal source that generates the pitch undulations.

Vibrato extraction begins with the analysis of pitch of singing voice, to identify its periodic variations around an average value. To compute pitch of the singing voice, we used the Cepstrum technique, as this method efficiently separates the excitation signal (enclosing vibration of the vocal folds) from the vocal tract [37]. We set the grid limit of vibrato extent to the maximum range of ± 1.5 semitone to extract the vibrato portion from the pitch contour. The part of pitch contour that lies within this semitone grid is identified as the vibrato. The identified pitch contour is then filtered using cascaded stages of high- and low-pass filters. We extracted two important parameters that define the vibrato completely—rate and extent. The vibrato rate is dynamically calculated from the magnitude spectrum of the filtered pitch contour by precisely finding the location of largest peak. The extent is determined from the maximum and minimum deviation of the pitch contour from F0 in each vibrato period. This generates 5-dimensional feature vector containing:

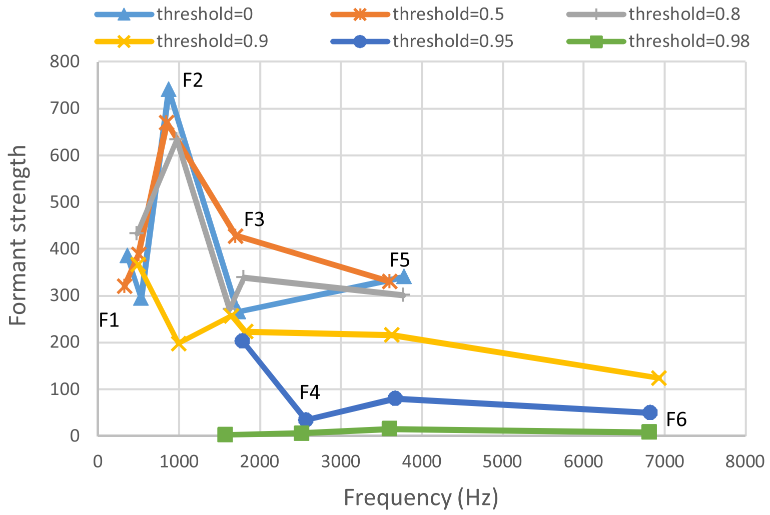


Fig. 2 Formants obtained for different condition on the root of LPC coefficients

vibrato rate and its peak value and vibrato extent, its maximum and minimum values. The vibrato estimation process is presented in Algorithm 2.

```
Algorithm 2: Vibrato estimation
Step1 – Similar to Step2 as mentioned in Algorithm 1.
Step2 – Find pitch ($F_o$) using Cepstrum technique
Step3 – Identify the sustained note (vibrato) of the singing segment from
        the pitch contour ($\hat{F}$) (Using ±1.5 semitones)
Step4 – Normalize the identified vibrato section
        $\widehat{F_N} = \frac{\hat{F}}{F_o} - mean(\frac{\hat{F}}{F_o})$
Step5 – Filter the normalized vibrato section (Eliminate frequencies below
        2Hz and above 10Hz)
Step6 – To calculate vibrato rate, compute spectrum of filtered vibrato
        section
        $Vibrato\ Rate = peak\{spectrum\}$
Step7 – For extent, compute max & min displacement of vibrato section
        $dmax_i = 1200 * |log_2(\widehat{F_N}(maxima(i)))|$
        $dmin_j = 1200 * |log_2(\widehat{F_N}(minima(j)))|$
        $extent = \frac{\sum_{i=1}^{I} dmax_i + \sum_{j=1}^{J} dmin_j}{I+J}$
    $i, j$ indicate the maxima's and minima's in the complete vibrato
    section
    $I, J$ indicate the total number of maxima and minima points.
```

3.3 Harmonic spectral envelope feature

In a singing voice, the periodic but abrupt opening and closing of the vocal folds produces a glottal flow waveform rich in harmonics. Soprano singers are high pitched singers and have harmonics which are widely spaced as compared to bass and tenor singers. The analysis of harmonic spectrum is useful in differentiating between low and high pitch singers. Harmonics and formant location together makeup the spectral shape of the sound. Harmonic spectral envelope (HSE) is a novel feature we have used for singer identification. We first obtained the spectral envelope of the singing segment using LPC. The harmonic spectral envelope is then constructed using the spectral envelope. We have described the harmonic spectral envelope as a function of harmonic vector [2F0, 3F0, 4F0, 5F0…]. The spectral envelope is now represented at the sample points of harmonic vector. The obtained envelope is termed as harmonic spectral envelope. The harmonic spectral envelope estimation process is presented in Algorithm 3.

```
Algorithm 3: Harmonic Spectral Envelope Estimation
Step1 – Similar to Step2 as mentioned in Algorithm 1.
Step2 – Find pitch ($F_o$) using Cepstrum technique
Step3 – Construct harmonic vector $HV = \{2F0, 3F0, 4F0, 5F0 \ldots \ldots \le \frac{Fs}{2}\}$
Step4 – Compute LPC coefficients (rts) of the input singing segment
        (LPC order =12).
Step5 – Obtain the spectral envelope
        $Y = 20 * log10\,\{|DFT(rts)|\}^{-1}$
Step6 – Represent spectral envelope as function of harmonic vector
        $Harmonic\ Spectral\ Envelope = Y(HV)$
```

As the harmonic spectral envelope exists only at the sampling instants of harmonics, it is relatively independent of interference of formants and formant variation.

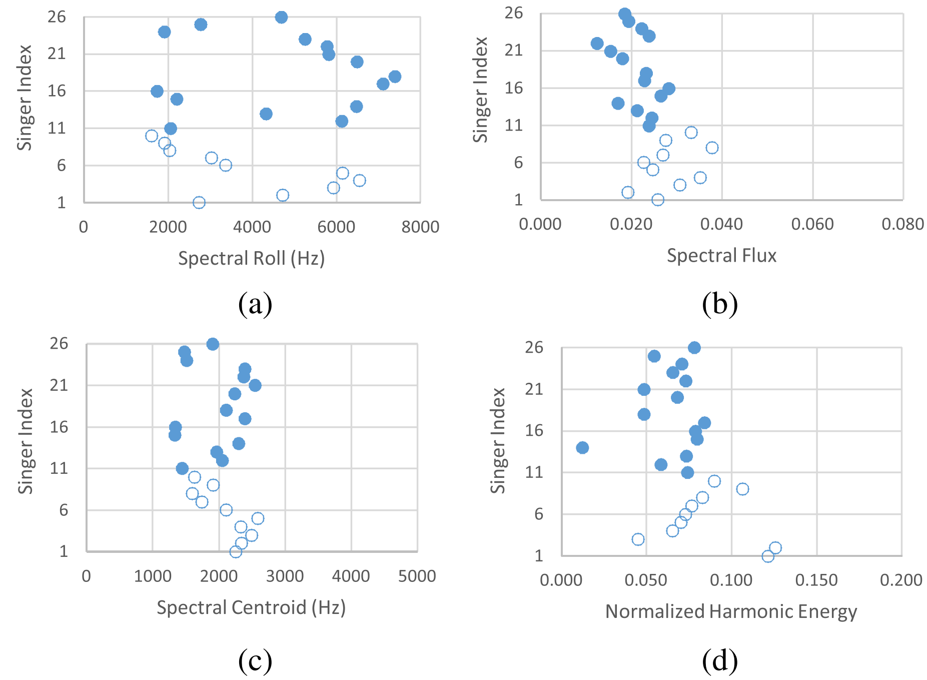


Fig. 3 a–d Analysis of spectral timbre features for each singer (blue circle indicates male singer; white circle indicates female singer)

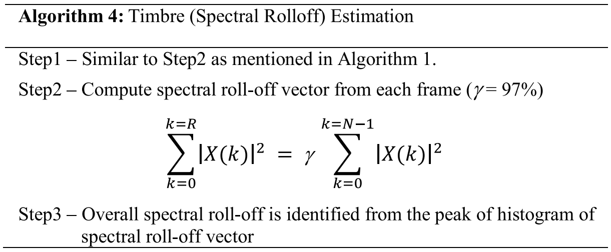
Also we observed consistency in its spectral shape from segment to segment for every singer. We explored this characteristic of the harmonic spectral envelope for the task of singer identification. To capture the spectral shape information, we considered 70-dimensional feature vector of harmonic spectral envelope.

3.4 Timbre feature

In context with music, timbre describes the characteristics sound of the instrument and its quality [38], while in context of singing voice, timbre helps to differentiate familiar voices. The different shape, size, and thickness of the vocal organs combine to create a particular timbre that personalizes the singer characteristics [39]. Timbre is mainly determined from the spectral components of the vocal organs that include the spectral envelope, the strength of harmonics, and how the sound varies with time. Typical timbral features obtained by capturing simple statistics of the spectra include spectral centroid, spectral roll-off, spectral flux, harmonic energy, spectral irregularity, spectral flatness, and spectral bandwidth [40].

We performed a detail analysis of spectral roll, flux, centroid, and harmonic energy, to identify the relevant timbre feature which could assist in describing the singer distinctively. These timbre features were analyzed for 26 singers. The plot of each feature displaying the mean value for every singer is shown in Fig. 3. The singer index 1 to 10 correspond to female singers, while 11 to 26 indicate male singers. The timbre feature suitable for the task of SID is one which exhibits distinct mean for each singer, and also the variance of the mean values is large among all singers. As clearly evident from Fig. 3, we observed such characteristics only for spectral roll feature (Fig. 3a) as compared to other timbre features. The analysis reveals that spectral roll is more suitable for SID task than the other timber features.

Moreover, the spectral roll-off and centroid are the features whose spectral attribute is a function of frequency, while the rest of spectral descriptors depend on the amplitude of spectral envelop. As the spectral energy distribution changes with the sound characteristics, the spectral flux and harmonic energy features demonstrate larger and random variations in their spectral values for a given singer. Such features are not suitable for SID. Whereas the spectral roll-off and centroid being function of frequency, they proved to be consistent among singers. But Fig. 3c reveals that the spectral centroid, i.e., the "center of mass" of the spectrum of almost all the singers, lies between 1500 and 2500 Hz. Based on these investigations, in the proposed work, we characterized timbre using only the spectral roll-off feature. We considered 11-dimendional feature vector computed from the peak values of spectral roll-off histogram. The spectral roll-off estimation process is presented in Algorithm 4.



```
Algorithm 4: Timbre (Spectral Rolloff) Estimation
Step1 – Similar to Step2 as mentioned in Algorithm 1.
Step2 – Compute spectral roll-off vector from each frame ($\gamma$ = 97%)
    $\sum_{k=0}^{k=R} |X(k)|^2 = \gamma \sum_{k=0}^{k=N-1} |X(k)|^2$
Step3 – Overall spectral roll-off is identified from the peak of histogram of
        spectral roll-off vector
```

3.5 Dimension reduction

As shown in Fig. 1, prior to the classifier, PCA is applied on the features extracted from the training samples. PCA is a powerful tool for analyzing data that helps to reduce the dimension, without much loss of information. Its goal is to extract the important information (i.e., capture most of the variance) from the data and represent it as a set of new orthogonal variables called principal components [41]. These principal components can be obtained by two methods covariance matrix and singular value decomposition. PCA actually extracts components equal to the number of observed variables being analyzed, but only the first few components are retained, interpreted, and used in subsequent analyses, as they account for meaningful amounts of variance [42]. In our system, we obtained 102-dimensional feature vector from formants (16-dimensional), vibrato (5-dimensional), harmonic spectral envelope (70-dimensional), and timbre (11-dimensional). PCA is applied on the input feature vector to obtain an information—concentrated feature vector of dimension 1 × 60.

3.6 Classifier—support vector machine

Support vector machines (SVM) are discriminative classifiers widely used for pattern recognition and machine learning. SVMs are basically linear two-class classifier that consists of objects labeled with one of two labels corresponding to the two classes [43]. For a given training data (X_i, Y_i) for $i = 1, 2, \ldots\ldots N$, with input vector $X_i \in \Re^d$ and label $Y_i \in \{-1, +1\}$, the decision boundary $f(X)$ must classify all the points correctly.

$$f(X_i) \begin{cases} \geq 0 \; Y_i = +1 \\ < 0 \; Y_i = -1 \end{cases}$$

The decision boundary is commonly referred to as the hyperplane, which separates data into parts, each part

Table 1 List of singers used in SID Experimentation

Singer index	Singer	Gender	Span of songs used in the database (year)	Language of the songs
1	Lata Mangeshkar	F	1965–1985	Hindi, Marathi
2	Kavita Krishnamurthy	F	1986–2012	Hindi
3	Sunidhi Chavan	F	2004–2015	Hindi, Telegu
4	Shreya Ghoshal	F	2008–2014	Hindi, Malayalam, Tamil, Kannada
5	Alka Yagnik	F	1988–2010	Hindi
6	Anuradha Paudwal	F	1985–2004	Hindi
7	Usha Mangeshkar	F	1964–1983	Hindi, Marathi
8	Suraiya	F	1947–1954	Hindi
9	Shamshad Begum	F	1949–1960	Hindi
10	Geeta Dutt	F	1958–1973	Hindi
11	Kishore Kumar	M	1962–1982	Hindi
12	Sonu Nigum	M	2000–2015	Hindi, Kannada
13	Amit Kumar	M	1979–2003	Hindi
14	Ankit Tiwari	M	2011–2016	Hindi
15	Mukush	M	1958–1970	Hindi
16	Rafi	M	1954–1979	Hindi
17	Adnan Sami	M	1998–2015	Hindi
18	Arijit Singh	M	2011–2016	Hindi
19	Kailash Kher	M	1995–2011	Hindi
20	Sukhwinder Singh	M	1998–2013	Hindi
21	Suresh Wadker	M	1982–1996	Hindi, Marathi
22	Kumar Sanu	M	1992–2004	Hindi
23	Manna Dey	M	1958–1981	Hindi
24	Mahendra Kapoor	M	1963–1974	Hindi
25	S P Balasubrahmanyam	M	1991–2015	Hindi, Telegu
26	K J Yesudas	M	1978–1990	Hindi

representing a class. This hyperplane could be linear or non-linear depending on the data. As linear decision functions are generally not rich enough for pattern separation, kernel functions can be applied [44]. The kernel function we used in our experimentation is Gaussian Radial Basis Function. The Gaussian radial basis function as defined in [45] is expressed in Eq. (2).

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{n} \alpha_i \exp\left\{-\frac{|x-x_i|^2}{\sigma^2}\right\}\right)$$

where x_i indicates the input data points and x indicates the testing data points; $|x - x_i|^2$ is the squared Euclidean distance between x_i and x; σ defines the width of the inner product kernel, i.e., area of influence the test data has over the data space; and α_i are scalar parameters. For extending the SVM to multiclass application, the practical alternative is to convert a two-class classifier to a multiclass. This is called one-vs-the-rest approach [43]. Here, we train the SVM for each singer using training samples from the database and classify the test sample according to the largest discriminant function value produced by the classifier.

4 Database

Variety of database exists for speech processing; however, researchers working in the field of music need to prepare their own database as per the requirement of one's research. Also, the corpus varies depending on the factors like the type of music, type of singers, ways of singing, type of recording, and mediums of data collection. The proposed work consist of in-house developed cappella database comprising songs of popular Indian playback singers who have contributed to Bollywood film industry over a span of more than 20 years. The songs are taken from commercially available CD recordings with sampling rate of 44.1 kHz. Table 1 shows the list of singers used in the SID experimentation.

Usually songs are accompanied with strong background music which is considered as the negative

Table 2 Effect of song duration on SID performance

	Song duration		
	3 s	6 s	9 s
SID accuracy (%)	79.73	82.02	86

influence affecting performance of singer identification system [46–48]. The features extracted from such songs do not solely represent the singing voice but a mixture of the singing voice and the background interference. Nwe and Li [5] in their research work have revealed that singer identification accuracy can be improved by using only vocal segments from the verse sections. To have a true representation of the estimated acoustic features and overcome the effect of the interfering background, we used the singing segments from the songs having no background interference.

Generally, a Bollywood Indian song structure has a definite pattern: prelude, the mukhda (face of the song), the interlude and the antaras! (Stanzas of the song). A song begins with prelude music followed by the mukhda, then continues with interlude music bridging the gap between two or three antaras of the song. The mukhda is repeated after every antara. The initial mukhda in large number of songs is usually sung with less musical accompaniment. We used this portion of the song for collecting cappella singing voice segments.

Table 1 includes the overall time span (years) of the songs included in the database for each singer. As the songs are collected from different Bollywood films, they are composed by diverse music composers, thereby changing the singing style of the singer. The database not only confines to songs in Hindi language but also includes songs in Marathi, Telegu, Kannada, and Malayalam. The purpose is to verify that singer characteristics are invariant to language of the lyrics. This led to the development of a complete non-homogenous database independent of type of the song and style of singing, indirectly testing the system for the so-called Album effect.

Table 3 Effect of multi-language songs on SID performance

Singer index	Language of the testing song lyrics	SID accuracy (%)	
		Segment Level	Song level
1	Marathi	66.66	75
3	Telegu	100	100
4	Malayalam, Tamil, Kannada	75	83.33
7	Marathi	66.66	76.66
12	Kannada	80	75
21	Marathi	65	72
25	Telegu	83.33	80
Average		76.66	80.28

A large cappella database is created as there is a huge availability and variety of Indian songs and singers. The proposed work includes analysis of 26 singers (10 female and 16 male singers) using 550 songs. Almost 20 songs per singer accounting to nearly 40 to 45 cappella segments (each of 3 s duration) per singer from different songs are annotated manually for training and testing of the system covering maximum variability in the singers' voice.

5 Experimental analysis and results

In this section, we first address the various SID investigations like effect of song duration and multi-language testing of singer songs and compare performance of classifiers for similar set of dataset and features. We would like to analyze the robustness of our SID system considering factors like singing style, album effect, and the language of singing lyrics. We also present the effectiveness of our SID system in identifying the cover song. In the last section, we compare the overall system performance with existing SID state-of-the-art approaches exploring Indian song and singers.

5.1 Effect of song duration on SID

The SID accuracy improves as song duration increases [13]. To analyze the effect of song duration on SID performance, we considered the songs of maximum 9-s song duration. Initially, the SID accuracy are calculated for each analyzing singing voice of duration 3 s. The song duration is then gradually increased, and the result of singer identification for different song duration is recorded as shown in Table 2. The analysis is the average accuracy of 26 singers. We computed the results for the combination of all the acoustic features (formants + HSE + vibrato + timbre). As expected, the SID accuracy increased as the song duration increased.

The SID accuracy is computed using Eq. (3)

$$\text{Accuracy } (\%) = \frac{\text{No.of correctly identified songs}}{\text{Total number of songs}} \tag{3}$$

We achieved an accuracy of 86% for a song duration of 9 s. We used fourfold cross validation approach to calculate the classification accuracy. In this approach, the entire samples from the database of each analyzing

Table 4 Performance of the classifiers for different PCA dimension

Classifier	PCA dimension	SID accuracy (%)
SVM	60	86
	10	65.11
GMM	60	41.29
	10	35.72

singer are grouped in four independent sets of nearly equal size, out of which three sets are used for training and one set is used for testing. The process is repeated four times till each set appears once as a test set.

5.2 Test of SID performance for songs of multiple language

The language of song lyrics makes a different impact on the listeners. In this section, we present the results of an experimentation performed to observe the impact of considering songs sung by the singers in multiple language on the SID system. The songs of Bollywood films are generally composed in Hindi language. Therefore, the initial challenge for the experimentation was to identify the multilingual singers from our Bollywood singer database. We identified seven singers from the database who have sung songs other than in Hindi language. A cappella database of such multilingual singers is created. We perform this experiment by training our system using only Hindi songs from the database. The system is then tested for multilingual songs of the test singer, i.e., songs other than in Hindi language. As observed in Table 3, singer with singer index 4 is tested for songs of Malayalam, Tamil, and Kannada languages.

To our knowledge, no other research work has attempted to test their SID system for multilingual songs. It is observed that multilingual singers often sing songs in a language which they do not understand. They sing these songs by interpreting the meaning of the lyrics and by the focus on the pronunciation of vowels and consonants. Bollywood singers singing South Indian songs especially Tamil, Kannada, and Telegu have to drastically change their singing style. They need to elaborate on the ragas, and there is possibility of losing the naturalness of voice under the burden of words and their emotion. This affects the acoustic features of the singer consciously or unconsciously. The identification accuracy shown in Table 3 is calculated at both segment and song level. An average accuracy of 80.28% is achieved at the song level in identifying the singers. One of the contributing factor for achieving this accuracy is the effect of amalgamation of multiple acoustic features. A closer observation revealed similar timbral characteristics among the songs of the multilingual singer. The result obtained clearly implies that our SID system is independent of language of the lyrics of the song and can effectively identify multilingual singers.

5.3 Classifier performance

Table 4 presents the comparison of SID results for two different classifiers considering different PCA feature vector dimension. We applied PCA on the input feature vector of dimension 1 × 102. The difference in the resultant accuracy reveals the impact of PCA feature vector dimension on the performance of SID system. Higher identification accuracy is obtained for PCA dimension of 60 as compared to at 10. It clearly indicates that reducing the dimension of the feature vector to smaller number of principal components accounts for loss of valuable information. This is because the input feature vector is constructed by combining four different acoustic features which are completely uncorrelated with one another. The PCA dimension of 10 is therefore not adequate to capture most of the essential variance of the input feature vector.

Using PCA components as the feature vector, we compared the performance of our SID system for two classifiers—the Gaussian mixture model (GMM) and support vector machine (SVM). GMM is another widely used classifier for MIR applications. In our experimentation, we build GMM model for each singer using the multiple features of each singer. The parameters of each GMM, the mixture weights, mean vectors, and covariance matrices are estimated from training samples of each singer. The covariance matrices of GMM used in this experimentation are diagonal. Mixtures of 16 Gaussians are trained with 100 iterations of the EM algorithm. For the given test singing segment, the posterior probabilities of all the singer classes are computed and the class with highest probability is identified as the class of the unknown test song. We observed that in comparison to Gaussian mixture model (GMM) classifier, the SVM provide better classification results. Using the fourfold cross-validation technique, we obtained 86% SID accuracy with SVM whereas GMM yielded to only 41.29%.

5.4 Robustness of SID system

The characteristics of Bollywood music are its diversity, addictiveness, and expressiveness. It is a blend of various singing styles, emotions, and rhythm. It is thus a challenging and essential aspect to check on the robustness of the SID system using the dynamically changing Bollywood corpus. In this section, we discuss on the robustness of our SID system.

5.4.1 Robustness to singing styles

Bollywood songs are the biggest fusion of music genre. The broad categories of Indian "art" styles are the Hindustani and the Carnatic. The Hindustani style of songs is observed mostly in the North and North West region of India while the Carnatic style is popularly governed in the Southern portion of India. The lyrics in Bollywood songs are framed describing the situations in the movies such as "heightening a situation, emphasizing a mood, and commenting on the theme and action" [49]. Subsequently to make songs suitable to such situations, Bollywood composers create a blend of the various song genres from classical, folk as well as western to

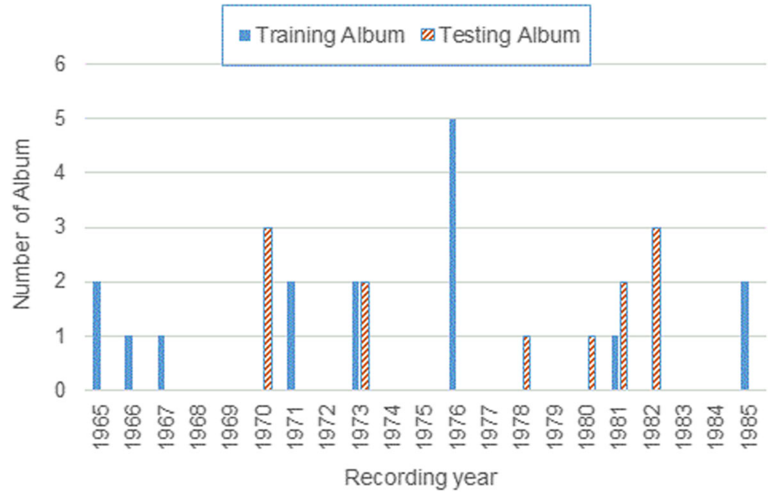


Fig. 4 Distribution of training and testing albums of a singer

formulate the new compositions. Also, the songs of Bollywood movies could be further categorized into romantic songs, sad songs, songs with Joy and Mirth, and prayer songs (Bhajans). Playback singers need to modulate their voice according to the category of the song, the situation in the movie and most important according to the actor or actresses to whom they are lending their voice.

Clearly, the large span of singer recordings used in our database as mentioned in Table 1 covers large number of movies and is a blender of variety of singing styles. The database includes singing segments without taking into consideration the singing style of the singer. The acoustic features extracted from such singing segments capture wide variations in the singer characteristics. Clearly, the basic characteristics of the Bollywood music itself make the database versatile and the system robust to the singing styles of the singer.

5.4.2 Robustness to album effect

In the western music, especially pop singing, it is observed that there is a close correlation of all the recordings in an album with the associated music or there is observed to be a consistency in the singing style of the singer within the album. Clearly, the systems using such corpus get trained to the album's style, rather than the singer's acoustic characteristics. This is called the “album effect” which greatly affects the singer identification process.

Table 5 Experimentation to find the cover singer

Singer singing the cover song (singer index)	Original singer of the song (singer index)	Accuracy in identifying the cover singer (%)	Identification of cover singer as the original singer of the song (%)
13	11	88	0
4	1	75	0

An album in Bollywood corresponds to songs of a single movie, with nearly five to six songs in each. The album effect is least observed with Bollywood movie song database. The major reasons are discussed below:

- All the songs in the initial era of song recording production were mostly composed by a single music composer. But the impact of the western music around 1970s on the Bollywood led to the development of multi-composer albums. Such albums lost the perception of commonness among the songs within a movie, producing songs with unique listening experience.
- Moreover as discussed in the earlier section, songs included in an album were composed with an intention to match the situation in the movie. Such restrictions resulted in songs with different singing style, emotion, and background music to be part of the same album.
- Furthermore, the songs in an album of Bollywood movies may not be necessarily sung by the same singer, but by different singers as selected by the music director.

Besides these significant factors, we have used database of 550 songs for the experimentation that includes albums which span songs of large recording period for each singer. The album effect is more pronounced when the songs are recorded within a relatively short time span, wherein it becomes the natural tendency of the composers to choose similar singing style and audio effects. The average recording span of the songs of all the singers in our database as mentioned in Table 1 is

Table 6 Performance comparison of the proposed system (*F* female singers, *M* male singers)

Acoustic feature	Experimental database	Song duration	Song level accuracy	Remarks
Formant + HSE + Vibrato + timbre	**Bollywood singers** **26 singers (10F, 16 M)** **550 songs**	**3–9 s**	**86%**	**Proposed work**
Method I: Carnatic interval cepstral coefficients Method II: MFCC [29]	Carnatic singers 6 singers 116 songs	–	Method I: 87% Method II: 56%	Lower singer dataset
Timbre [30]	North Indian singers 10 singers	5 s	58.33%	Lower singer dataset
Timbre + pitch + MFCC + LPC [31]	6 singers (2F, 4 M) 300 songs	180 s	81%	Audio portion taken from video songs. Lower singer dataset
MFCC + spectral features [32]	Tamil singers 3 singers (2F, 1 M) 300 video songs	10 s	95%	Audio portion taken from video songs. Lower singer dataset
Cepstral mean subtracted MFCC [34]	Bollywood singers 20 singers (6F, 14 M) 500 songs	60–166 s	84.5%	Longer Song duration
Perceptual linear prediction [50]	Tamil singers 10 singers (5F, 5 M) 200 songs	20 s	55.56%	Lower singer dataset

approximately 17 years. This large span of recording period clearly indicates the songs used in our experimentation are taken from wide span of albums. This distribution of albums is shown in Fig. 4 for one of the singer from our database (Singer 1). Both the training and testing albums are represented with their recording year. As shown in Fig. 4, more than 80% of the albums used in training and testing of the classifier are mutually exclusive. The cross validation process further shuffles the training and testing albums for each testing iteration of the singer songs. The system using such characteristics of the database is robust to the album effect.

5.4.3 Robustness in identifying cover songs

Singer adopts the singing style to suit the lyrics of the song. Amateur singers imitate the singing style of the original singer and are to some extent successful in creating a perpetual impression of the original singer. Such recordings of original singers by other artists are known as cover version or cover songs. The SID system should be robust in identifying such cover version of the songs.

Table 7 Performance evaluation of existing Indian SID systems (results are tested using our database of 26 singers with maximum song duration of 9 s)

Method	SID accuracy (%)
Formant + HSE + Vibrato + Timbre	**86**
Method II: MFCC [29]	26.50
Timbre [30]	25.05
Timbre + pitch + MFCC + LPC [31]	23.21
MFCC + spectral features [32]	38.38
Cepstral mean subtracted MFCC [34]	28.49
Perceptual linear prediction [50]	45.69

To examine the robustness of our system in this direction, we collected cover songs and created an additional database of such songs. Table 5 lists the singers of the cover songs along with the original singer of the cover songs. To have a systematic analysis in identifying the singer of the cover song, we collected cover songs sung by singers listed in our database. Furthermore, the original singers of the cover songs are also the singers included in our database. To further increase the complexity, we included the original songs of the cover version in the training set of the original singer. The SID system is tested for the cover songs, and the singer identification accuracy of these cover songs is presented in Table 5. The results show that our SID system correctly identifies the cover singer but none of the singers of the cover songs are identified as the original singer of the cover song. This clearly indicates that the classifier of our system captures the characteristics of the singer, rather than the lyrics and structure of song. Results show the proposed SID system is robust in identifying the cover singers and is independent of singing lyrics.

5.5 SID performance comparison

Table 6 shows a detailed comparison of the proposed work with other existing research works that have experimented on Indian songs and singers.

For comparison, the performance parameters considered are volume of dataset and song duration. It is clearly evident that there has not been much research on singer identification of Indian singers. Moreover, many of them are based on single acoustic feature and have experimented on lower singer dataset. The other important performance factor is the song duration considered for SID. The song duration we used for the experimentation is least and efficient in identifying singers from large dataset, as compared to other research work using longer song duration for lower volume of singer dataset. We have claimed identification accuracy of 86% using song duration of maximum 9 s. Also, the analysis presented in Table 2 supports that SID accuracy can be further improved as the song duration increases.

To compare our SID system, we further evaluated the performance of these existing SID methods using our dataset. To ensure fair evaluation, we maintained the same settings (like feature extraction approach and dimension of feature vector) as mentioned in each research work. The intention was to record the performance of these methods for larger singer dataset and limited song duration. Table 7 presents the performance evaluation of six existing methods (as mentioned in Table 6) with our SID system.

The results in Table 7 clearly revealed that the accuracy of all the SID techniques is lower than the result obtained by us when evaluated on our dataset. The factors that affected the accuracy of these methods are the increased singer count and lowered song duration, while the appropriate selection of acoustic features in our system accounted for higher accuracy. The comparison of the result with the mentioned researches clearly indicates that our system is more efficient and robust for MIR applications with reference to singer identification of Indian playback singers. We strongly claim that ours is the first SID system considering the four features, viz, formants, harmonic spectral envelope, vibrato, and timbre, which is tested on large number of Indian playback singers and songs.

Few other observations from other research work are mentioned here:

- The training and testing corpus duration are of different length [11, 33].
- Consideration of association between instrument configuration used and the songs performed by the singer [28, 51], i.e., singers tend to work with particular set of backing instruments.
- Create or record the dataset within a restricted environment with a focus towards the acoustic feature under investigation [27, 52].

Our presented work does not confine to any of these hypotheses and proves its superiority over existing approaches.

6 Conclusion

The research work has presented a framework that primes to robust identification of the singers considering multiple acoustic features of singing voice. We examined the feasibility of the system on the database of pure cappella samples of Bollywood playback singers and measured the performance of these acoustic features in describing the singer characteristics. The experimental work revealed that use of multiple features can effectively capture the singer characteristics from a heterogeneous database.

Our experimentation found that the higher formants personalize the voice characteristics. Further, it demonstrated that each singer has a specific pattern of the spectral harmonic curve. The distribution of vibrato parameters—rate and extent—proves vibrato as one of the informative cues for singer identification task. We explored the various spectral features of timbre and found spectral roll-off correlating to singer characteristics by providing distinct band of roll-off frequencies for each singer.

For the first time, we have experimentally demonstrated that SID system can assist in identifying cover-version songs which consist of re-recorded famous songs by singers, having similar singing voice perception as the original singer. Also, the effectiveness of SID system in identifying the singer when it is tested for songs in different languages is also demonstrated. This is due to the fact that the proposed system is independent of melody and music structure accompaniment of the original song. The singer identification system is also robust to album effect, as the system is tested for variety of singer songs with different singing style, period of recording, and language of the lyrics and includes songs composed by diverse music composers.

In the future, we plan to extend the work to other MIR applications like singer tracking in case of duet songs and separation of singing voice from accompanied music and analyze the variations in the acoustic features considering songs over the performance period of a singer.

Abbreviations

GMM: Gaussian mixture model; HSE: Harmonic spectral envelope; LPC: Linear predictive coding; MFCC: Mel frequency cepstral coefficients; MIR: Music information retrieval; PCA: Principal component analysis; SID: Singer identification; SVM: Support vector machine

Authors' contributions

DL has performed the experimental work and authored the paper. The entire work has been under the supervision and guidance of SS. Both authors read and approved the final manuscript.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

"Not applicable".

Consent for publication

"Not applicable".

Competing interests

The authors declare that they have no competing interests.

Author details
[1]D.K.T.E's Textile & Engineering Institute, Ichalkaranji, India. [2]Walchand College of Engineering, Sangli, India.

Received: 31 October 2018 Accepted: 26 May 2019
Published online: 17 June 2019

References

1. D.Y. Loni, S. Subbaraman, Extracting acoustic features of singing voice for various applications related to MIR: A review. Proc. Int. Conf. Adv. Signal Process. Commun., 66–71 (2013) Washington, DC: ACEEE. DOI: 03.LSCS.2013.3.520
2. J. Sundberg, The acoustics of the singing voice. Sci. Am. **236**, 82–91 (1977)
3. N. Migita, M. Morise, T. Nishiura, in *Proceedings of 20th International Congress on Acoustics*. A study of vibrato features to control singing voices (2010), pp. 1–4
4. T. Saitou, M. Unoki, M. Akagi, Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis. Speech Comm., 405–417 (2005). https://doi.org/10.1016/j.specom.2005.01.010
5. T.L. Nwe, H. Li, Exploring vibrato-motivated acoustic features for singer identification. IEEE Trans. Audio Speech Lang. Process. **15**(2), 519–530 (2007). https://doi.org/10.1109/TASL.2006.876756
6. A. Butler, V.R. Lind, K. Van Weelden, Research on the aging voice: Strategies and techniques for healthy choral singing. J. Phenom. Sing. **1**, 42–50 (2001)
7. J. Shen, J. Shepherd, A.H.H. Ngu, Towards effective content-based music retrieval with multiple acoustic feature combination. IEEE Trans. Multimedia **8**(6), 1179–1189 (2006). https://doi.org/10.1109/TMM.2006.884618
8. A. Behl, M. Choudhury, in *9th International Conference on Natural Language Processing Macmillan Publishers, India*. A Corpus linguistic study of Bollywood song lyrics in the framework of complex network theory (2011)
9. A. Morcom, *Hindi Film Songs and the Cinema* (Burlington, VT :Ashgate, Aldershot, 2007)
10. B. Whitman, G. Flake, S. Lawrence, in *Proceedings IEEE Workshop on Neural Networks for Signal Processing, Falmouth, Massachusetts*. Artist detection in music with Minnowmatch (2001), pp. 559–568. https://doi.org/10.1109/NNSP.2001.943160
11. T. Zhang, *Automatic Singer Identification* (Proceedings of ICME, Baltimore, 2003)
12. W.-H. Tsai, H.-C. Lee, Singer identification based on spoken data in voice characterization. IEEE Trans. Audio Speech Lang. Process. **20**(8), 2291–2300 (2012)
13. W.-H. Tsai, H.-M. Wang, Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals. IEEE Trans. Audio Speech Lang. Process. **14**(1), 330–341 (2006). https://doi.org/10.1109/TSA.2005.854091
14. M. Lagrange, A. Ozerov, E. Vincent, in *Proc. of the 13th Int. Society for Music Information Retrieval Conference (ISMIR), Porto*. Robust singer identification in polyphonic music using melody enhancement and uncertainty-based learning (2012), pp. 595–600
15. A. Mesaros, T. Virtanen, A. Klapuri, in *Proc. 8th Int. Conf. Music Inf. Retrieval*. Singer identification in polyphonic music using vocal separation and pattern recognition methods (2007), pp. 375–378
16. A. Holzapfel, Y. Stylianou, *Singer Identification in Rembetiko Music* (Sound and Music Computing Conference (SMC), Lefkada, 2007)
17. S. Shirali-Shahreza, H. Abolhassani, M. Shirali-Shahreza, Fast and scalable system for automatic artist identification. IEEE Trans. Consum. Electron., 1731–1737 (2009). https://doi.org/10.1109/TCE.2009.5278049
18. Y.E. Kim, B. Whitman, in *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR)*. Singer identification in popular music recordings using voice coding features (Paris, 2002), pp. 164–169
19. M. Mauch, H. Fujihara, K. Yoshii, M. Goto, in *12th Int. Proc. on Music Information Retrieval*. Timbre and melody features for the recognition of vocal activity and instrumental solos in polyphonic music (2011), pp. 233–238
20. Andersen, in *Cognitive Inf. Processing (CIP) 4th Int. Workshop*. Using the Echo Nest's automatically extracted music features for a musicological purpose (2014), pp. 1–6
21. T. Saitou, M. Goto, in *Proc. of International Speech Communication Association*. Acoustic and perceptual effects of vocal training in amateur male singing (2009), pp. 832–835
22. I. Arroabarren, M. Zivanovic, J. Bretos, A. Ezcurra, A. Carlosena, Measurement of vibrato in lyric singers. IEEE Trans. Instrum. Meas. **51**, 660–665 (2002). https://doi.org/10.1109/TIM.2002.803082
23. M. Mellody, F. Herseth, G.H. Wakefield, Modal distribution analysis, synthesis, and perception of a soprano's sung vowels. J. Voice **15**(4), 469–482 (2001). https://doi.org/10.1016/S0892-1997(01)00047-9
24. Peter Desain, Henkjan Honing, Rinus Aarts and Renee Timmers, rhythmic aspects of vibrato. Proceedings of Rhythm Perception and Production Workshop, vol.34, pp. 203–216 (1999)
25. T. Nakano, M. Goto, Y. Hiraga, in *Proc. of International Speech Communication Association*. An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features (2006), pp. 1706–1709
26. A. Bonjyotsna, M. Bhuyan, Analytical study of vocal vibrato and mordent of Indian popular singers. J. Voice **30**(6), 764.e11–764.e22 (2015). https://doi.org/10.1016/j.jvoice.2015.10.010
27. A. Mesaros, S. Moldovan, in *IEEE International Conference on Automation, Quality and Testing, Robotics*. Method for singing voice identification using energy coefficients as features (2006), pp. 161–166. https://doi.org/10.1109/AQTR.2006.254623
28. Jialie Shen, John Shepherd, Bin Cui, Kian-Lee Tan, A Novel Framework for Efficient Automated Singer Identification in Large Music Databases. ACM Transactions on Information Systems, 27 3, 1–31 (2009) DOI https://doi.org/10.1145/1508850.1508856
29. R. Sridhar, T.V. Geetha, in *IEEE conference on Computer Science*. Music information retrieval of Carnatic songs based on Carnatic music singer identification (2008), pp. 407–411
30. S. H. Deshmukh, S.G. Bhirud, North Indian classical Music's singer identification by timbre recognition using MIR toolbox. Int. J. Comput. Appl. **91**(4), 1–4 (2014). https://doi.org/10.5120/15866-4804
31. T. Ratanpara, N. Patel, *Singer identification using perpetual features and cepstral coefficients of an audio signal from Indian video songs. EURASIP Journal on Audio, Speech, and Music Processing* (2015). https://doi.org/10.1186/s13636-015-0062-9
32. S. Metilda Florence, S. Mohan, A novel approach to identify a singer in a video song using spectral and cepstral features. J. Chem. Pharm. Sci. **10**(1), 462–465 (2017)
33. W. Cai, Q. Li, X. Guan, in *Proc. of the International Conference on Natural Computation, Shanghai*. Automatic singer identification based on auditory features (2011), pp. 1624–1628. https://doi.org/10.1109/ICNC.2011.6022500
34. H.A. Patil, P.G. Radadia, T.K. Basu, in *IEEE International Conference on Asian Language Processing, Hanoi*. Combining evidences from Mel cepstral features and cepstral mean subtracted features for singer identification (2012), pp. 145–148. https://doi.org/10.1109/IALP.2012.33
35. D. O'Shaughnessy, *Speech Communication: Human and Machine- Addison-Wesley Series in Electrical Engineering* (1987)
36. Karyn O'Connor, Singing – Singing – Singing with An 'Open Throat': Vocal Tract Shaping Formants and Tone. Singwise.com.,N.p,.2016. Wb. 2016
37. D.Y.Loni, Dr. S.S.Subaraman, Formant Estimation of Speech and Singing Voice by Combining Wavelet with LPC and Cepstrum Techniques. IEEE International Conference on Industrial & Information Systems, Gwalior,pp.1–7 (2014) https://doi.org/10.1109/ICIINFS.2014.7036530
38. N.W.E. Tin Lay, H. Li, in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process*. On Fusion of Timbre-Motivated Features for Singing Voice Detection and Singer Identification (2008), pp. 2225–2228. https://doi.org/10.1109/ICASSP.2008.4518087
39. A. Mesaros, *Singing Voice Recognition for Music Information Retrieval. Tampere University of Technology*, vol 1064 (Tampere University of Technology. Publication, 2012), p. 77
40. Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang, A Survey of Audio-Based Music Classification and Annotation. IEEE Transactions on Multimedia, vol. 13, 2, pp. 303–319 (2011) https://doi.org/10.1109/TMM.2010.2098858
41. H. Adbi, L.J. Williams, Principal component analysis. Comput. Stat. **2**, 433 (2010)
42. A. Tharwat, Principal component analysis—A tutorial. Int. J. Appl. Pattern Recognit. **3**(3), 197–240 (2016)
43. A. Ben-Hur, J. Weston, A. User, S guide to support vector machines. Methods Mol. Biol. **609**, 223–239 (2010). https://doi.org/10.1007/978-1-60327-241-4_13.
44. J. Wang, P. Neskovic, L.N. Cooper, in *Proceedings of the First International Conference on Advances in Natural Computation* - Volume

Part I. ICNC'05. Training data selection for support vector machines (Springer-Verlag, Changsha, 2005), pp. 554–564. https://doi.org/10.1007/11539087_71

45. Jesus Olivares-Mercado, Gualberto Aguilar, Karina Toscano-Medina, Mariko Nakano and Hector Perez Meana, GMM vs SVM for Face Recognition and Face verification . Reviews, Refinements and New Ideas in Face Recognition. Dr. Peter Corcoran (Ed.) (2011)
46. Yipeng Li and DeLiang Wang, Separation of Singing Voice From Music Accompaniment for Monaural Recordings. IEEE Trans. on Audio, Speech, and Language Processing, vol. 15, 4, pp. 1475–1487 (2007) https://doi.org/10.1109/TASL.2006.889789
47. Chao-Ling Hsu, DeLiang Wang, Jyh-Shing Roger Jang, and Ke Hu, A Tandem Algorithm for Singing Pitch Extraction and Voice Separation From Music Accompaniment. IEEE Trans. on Audio, Speech, and Language Processing, 20, 5, pp. 1482–1491 (2012) https://doi.org/10.1109/TASL.2011.2182510
48. T.L. Nwe, Y. Wang, in *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR '04)*. Automatic Detection of Vocal Segments in Popular Songs (Barcelona, 2004), pp. 138–145
49. R.B. Mehta, R. Pandharipande, *Bollywood and Globalization: Indian Popular Cinema, Nation, and Diaspora* (Anthem Press, 2010), p. 36 ISBN 978–1–84331-833-0. https://www.jstor.org/stable/j.ctt1gxp6bs
50. D. Dharini, A. Revathy, in *International Conference on Communication and Signal Processing*. Singer identification using clustering algorithm (India, 2014), pp. 1927–1931. https://doi.org/10.1109/ICCSP.2014.6950180
51. N.C. Maddage, C. Xu, Y. Wang, in *Proc. Int. Conf. Pattern Recognition*. Singer identification based on vocal and instrumental model (Cambridge, 2004), pp. 375–378
52. M.A. Bartsch, G.H. Wakefield, Singing voice identification using spectral envelope estimation. IEEE Trans. Speech Audio Process. **12**(2), 100–109 (2004). https://doi.org/10.1109/TSA.2003.822637

7 Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.