Singing Voice Analysis for Singer Identification Using Vibrato Features

Deepali Y Loni* and Dr Shaila Subbaraman**
*Department of Electronics, Textile & Engineering Institute Ichalkaranji, India
deepaliloni@rediffmail.com
**Ex-Academic Dean, Walchand College of Engineering Sangli, India
shailasubbaraman@yahoo.co.in

Abstract: Vibrato is one of the key acoustic feature of singing voice. This work investigates the measurement of vibrato by extracting it from natural cappella section of singing voice considering variety of singing styles and tones. The algorithm performs detection, extraction and analysis of vibrato parameters; rate and extent. These vibrato parameters are explored for singer identification and is validated on a database of 11 singers (5 female and 6 male singers) containing more than 50 cappella segments from several songs of each singer derived from commercially available CD recordings. A maximum accuracy of 62% is achieved in identifying the singers. This indicates the significance of vibrato as an important characteristic of singing voice.

Keywords: Vibrato, cappella, pitch contour, singer, confusion matrix.

Introduction

One of the widely used quality measuring benchmark of singing is vibrato. Vibrato is the skill acquired by the singers after many years of extensive vocal training. The vibrato adds certain naturalness to the singing voice and it is a very specific characteristic of the singing voice of a singer [1]. It is thus considered as one of the distinct factor that differentiates singing from plain speech. Vibrato is defined as periodic variations of the fundamental frequency of the singing voice around an average value.

"Vocal Vibrato" is also said to be the ornament of singing voice. Singers use vibrato to enhance their expressiveness and to place emphasis on significant words or phrases of a musical piece [2]. Major of the research work in this direction has revealed that trained singers have dominant presence of vibrato in the singing segments as compared to untrained singers [3]. Investigations have also revealed that vocal training improves vibrato as well as other acoustic features like formant and perceived quality of the singing voice [4].

The natural production of vibrato sound is the result of rhythmic variation of larynx muscles in response to sub-glottic pressure [5] which indirectly is singer specific characteristic and can be explored for singer identity. Edward [6] found that not all singers present the same quality of vibrato in terms of deviation of vibrato confidence. The results presented in [7] revealed the variations in the vibrato of the professional singers and also indicated that when a singer imitates another singer, the vibrato features are consciously controlled. The effectiveness of vibrato-motivated feature; octave frequency cepstral coefficients (using triangular, parabolic and cascaded subband filters) is explored by Tin Lay et. al. [8] in order to identify singers of popular music.

The proposed work explores vibrato characteristics of different singers. Major of the earlier studies extract vibrato by making the singers sing the same tone in rather slow tempo to produce large number of vibrato cycles. In contrast to these set conditions, the proposed investigation analyzes the vibrato information of natural singing voice by identifying the vibrato sections present in different segments of a song. The vibrato parameters; rate and extent are computed to analyze and identify the singer.

Extracting Vibrato Sections from Singing Voice

The most important characteristics that define the vibrato acoustic are the vibrato frequency (measured in Hertz) and the vibrato extension (measured in semitones). The vibrato frequency is a measurement of the speed of pitch variation; commonly called as vibrato rate. The vibrato extension also known as vibrato extent is the measurement of vibrato depth. A typical vibrato segment extracted from a singing voice is as shown in Fig. 1.

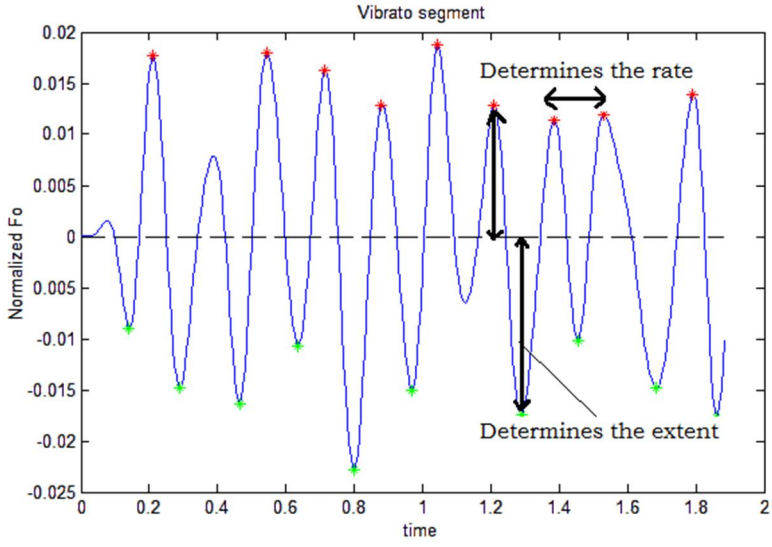


Figure 1. A vibrato segment

Typically, the vibrato occurs on sustained notes as it describes a slow continuous variation of frequency over its duration. A stable pitch contour corresponds to a note and melody of the song decides the note change. Complete information about rate, extent and shape of each period of pitch vibrato can be extracted from the pitch contour. Therefore initially the pitch contour is the major acoustic factor to deal with. Pitch contours represent time and frequency continuous sequences of pitch candidates [9] of each segment of singing voice under consideration. The average pitch of the pitch contour is considered as the current note of singing segment. The plot of pitch contour extracted from a singing segment is shown in Fig. 2 wherein the section containing the vibrato information is identified.

The semitone grid limit is set to ±1.5 [10] and the part of pitch contour that lies within this semitone grid is identified as the vibrato section. The identified vibrato section in Fig. 2 has the pitch contour which vibrates between 525Hz and 570Hz.

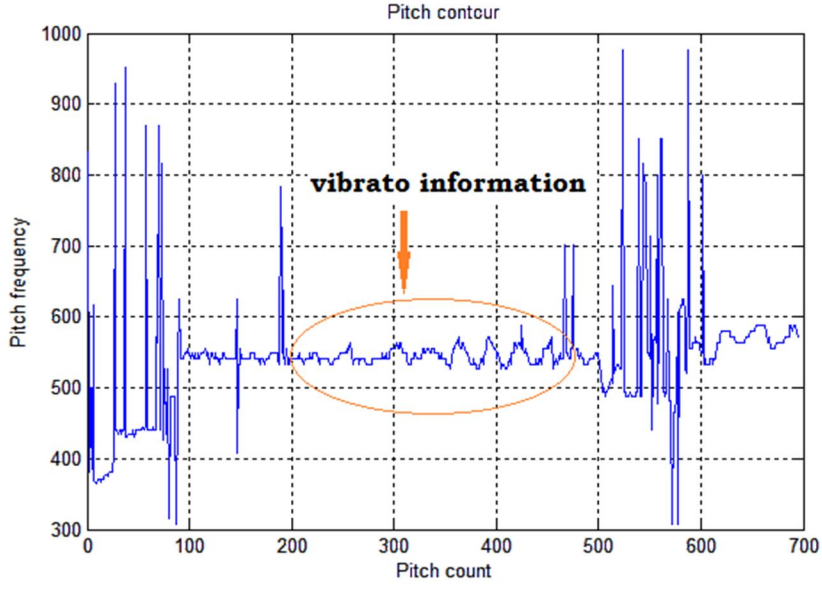


Figure2. Presence of vibrato in a pitch contour

Vibrato Parameters

Other than rate and extent vibrato is also described by regularity and waveform. Regularity and waveform describe the appearance of vibrato and are of lesser perceptual relevance in labelling singer specific characteristics. The presented work focus on rate and extent as they represent the key vibrato parameters to compare different singers.

Measuring the Vibrato Rate

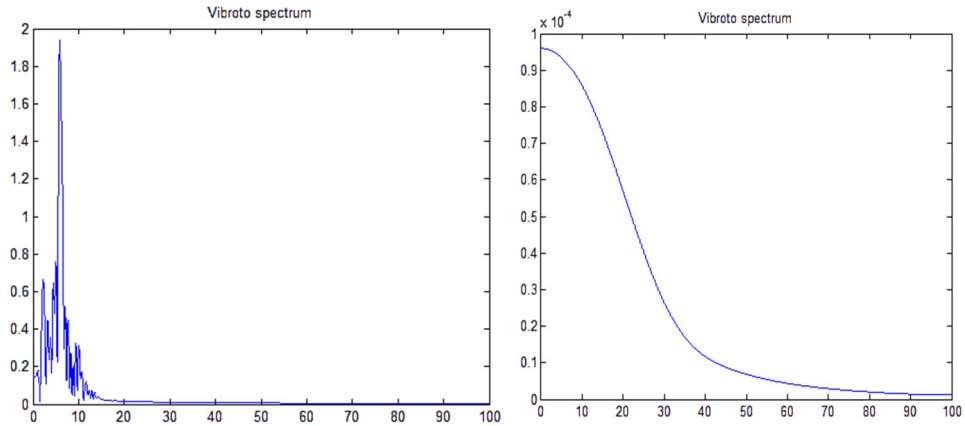
The vibrato rate is the number of vibratory cycles per second expressed in Hertz and is extracted from a vibrating pitch contour. To have accurate analysis and estimation of pitch contour, the research work focuses on cappella regions of singing voice.

The analysis length of cappella singing considered in the pitch extraction process is three seconds. Recordings shorter than three seconds are discarded. With a frame length of 25msec and overlap of 5msec, 821frames are obtained. Using the technique of Cepstrum and applying center clipping autocorrelation method, pitch of singing voice of various frames are computed. A pitch histogram is then figured from the pitch extracted from these frames. Finally the pitch of the cappella (F_o) is the average of the pitch values that lie in the highest bin of histogram. The computed pitches are assembled in time to form the pitch contour as shown in Fig. 2 and as discussed the vibrato section is extracted.

Let $\hat{F}$represent the pitch contour of the extracted vibrato section. $\hat{F}$is first normalized with respect to its average pitch value (F_o). This normalization helps to better capture pitch oscillations. The mean of the normalized pitch contour is then eliminated so that the pitch contour is centered on zero and this assists in obtaining smooth deviation of pitch contour about the mean.

$$\widehat{F_N} = \frac{\hat{F}}{F_o} - mean\left(\frac{\hat{F}}{F_o}\right)$$

The zero mean normalized pitch contour ($\widehat{F_N}$) is then passed through a cascaded stages of high and lowpass filters. The components in $\widehat{F_N}$ below 2Hz (outside the interest band) are eliminated using sixth-order highpass type-II Chebyshev filter with 30 dB attenuation. The resulting pitch contour curve is then filtered using lowpass sixteenth-order type-II Chebyshev filter with a cutoff frequency of 10 Hz. The various fluctuations (overshoot, vibrato, and preparation) [11] observed in the pitch contour are eliminated in the filtering process. The spectrum of filtered contour is then computed. As the range of vibrato rate is narrow, the N-point DFT is set large intentionally to obtain high frequency resolution of vibrato rate. The vibrato rate is dynamically calculated by finding the precise localization of the maxima of the magnitude spectrum as seen in Fig. 3(a). In these experiments it is observed that, vibrato is absent in some segments of singing voice. As is evident from Fig. 3(b) there is absence of vibrato data in the segment under consideration.



a) Vibrato rate = 5.85 b) No vibrato exists

Figure3. The vibrato spectrum

Measuring the Vibrato Extent

The vibrato extent describes how far the pitch rises and falls from the average pitch of a vibrato segment under consideration [12]. The extent of vibrato is maintained at ±1.5 semitone. The maxima and minima of the pitch contour are identified and their displacements from the average pitch value is computed in each vibrato period. Maxima less than 0.1s apart are eliminated to reduce the effect of noise fluctuations.

The challenging task while measuring extent is to find the true maxima and minima points of the pitch contour. The peaks obtained from the filtered pitch contour may not closely indicate the true maxima and minima of the original pitch contour as revealed in Fig. 4.

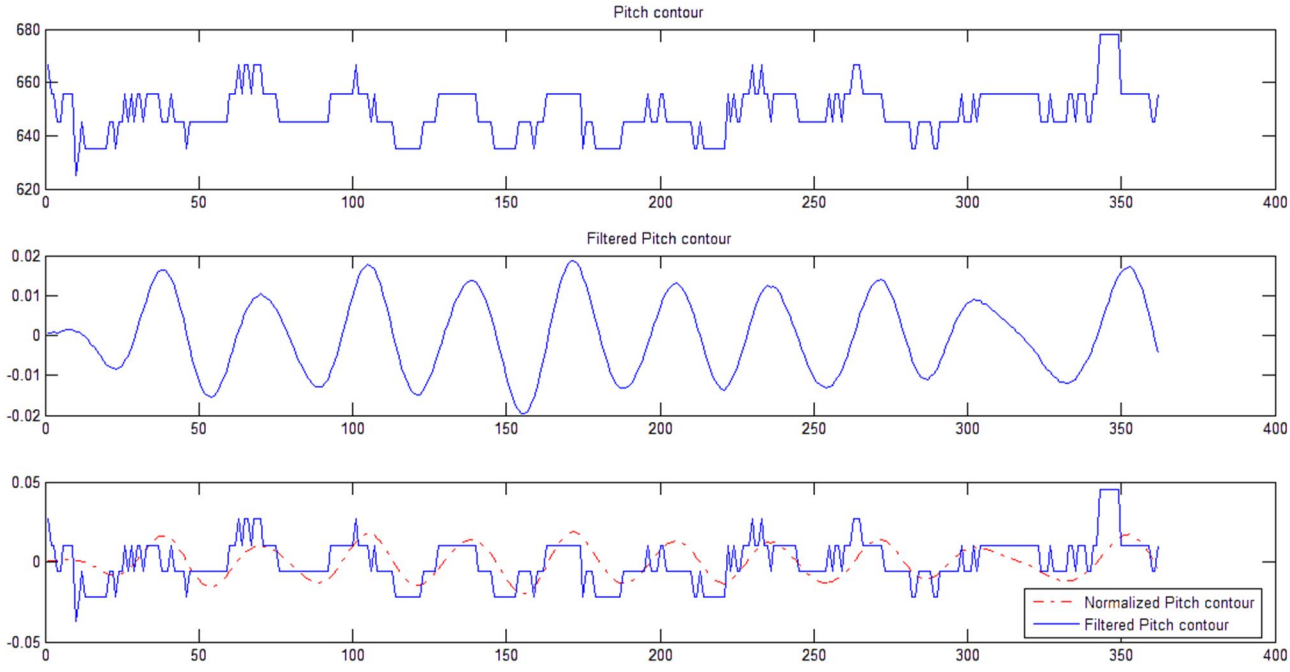


Figure 4. Maxima and minima points of pitch contour

The precision associated with finding the peaks is improved by proper selection of filter parameters and design of the peak finding algorithm. The extent is estimated as the mean of the maximum displacements and minimum displacements from the average pitch value observed in each vibrato period. The maximum displacement (*dmax*) and minimum displacement (*dmin*) is calculated by

$$dmax_i = 1200 * \left|log_2(\widehat{F_N}(maxima(i)))\right|$$

$$dmin_j = 1200 * \left|log_2(\widehat{F_N}(minima(j)))\right|$$

Where $\widehat{F_N}$ is zero mean normalized pitch contour, $maxima(i)$ & $minima(j)$ are the maxima and minima of the filtered pitch contour and i, j indicate the maxima's and minima's in the complete vibrato section. The logarithmic unit cent is used to express vibrato extent which is a relative unit widely used in musical domain. The overall extent is

$$extent = \frac{\sum_{i=1}^{I} dmax_i + \sum_{j=1}^{J} dmin_j}{I + J}$$

Here, I, J indicate the total number of maxima and minima points.

Analysis of Vibrato Parameters

The vibrato parameters viz. rate and extent are extracted from the database of cappella singing voice of famous singers from Indian film songs. The songs are extracted from commercial CD recordings with sampling rate of 44.1 kHz, down-sampled to 40 kHz. The vibrato extracted from these samples in most of the cases exceed more than four vibrato cycles, wherein the rate and extent can be precisely computed. But certain samples have short vibrato parts resulting in a single vibrato cycle. However complete vibrato information can be acquired even from one vibrato cycle, which has also been reported by David M. Howard et.al. [13]. The proposed work includes analysis of eleven singers covering maximum variability in the singers' voice.

Analysis of Vibrato Rate

It is claimed that the vibrato rate lies within the values 2 Hz to 8Hz [6] but the proposed work has extended the range from 2Hz to 10Hz to cover wide range of expected values. In the following six plots of Fig. 5, the vibrato rate histograms of few singers are plotted (three female –F1, F2, F3 and three male- M1, M2, M3).

Fig. 5 indicates the distribution of vibrato rate of different singers. As clearly evident the vibrato rate spreads in the entire band from 2 to 10 Hz, but still each singer demonstrates occurrence of his/her own dominant area of vibrato rate occurrence. Singer F1 shows more occurrence of rate in a range from 5 to 6.3 Hz whereas Singer F3 has this rate more prominently in the range 3 to 4.5 Hz. The prominent area of Singer M2 happens to be mostly flat area of Singer M1. The pattern of vibrato rate histogram reveals the individualities of singers. The average vibrato rates of singers are 5.56 Hz, 5.47 Hz, 5.10 Hz for female singers F1, F2 and F3 respectively whereas 5.61 Hz, 5.59 Hz, 5.71 Hz for male singers M1, M2 and M3 respectively. The average vibrato rate of male singers tend to be slightly higher than the female singers. The maximum recorded vibrato rate is found to be 9.32 Hz for a male singer from the database of singers used for experimentation.

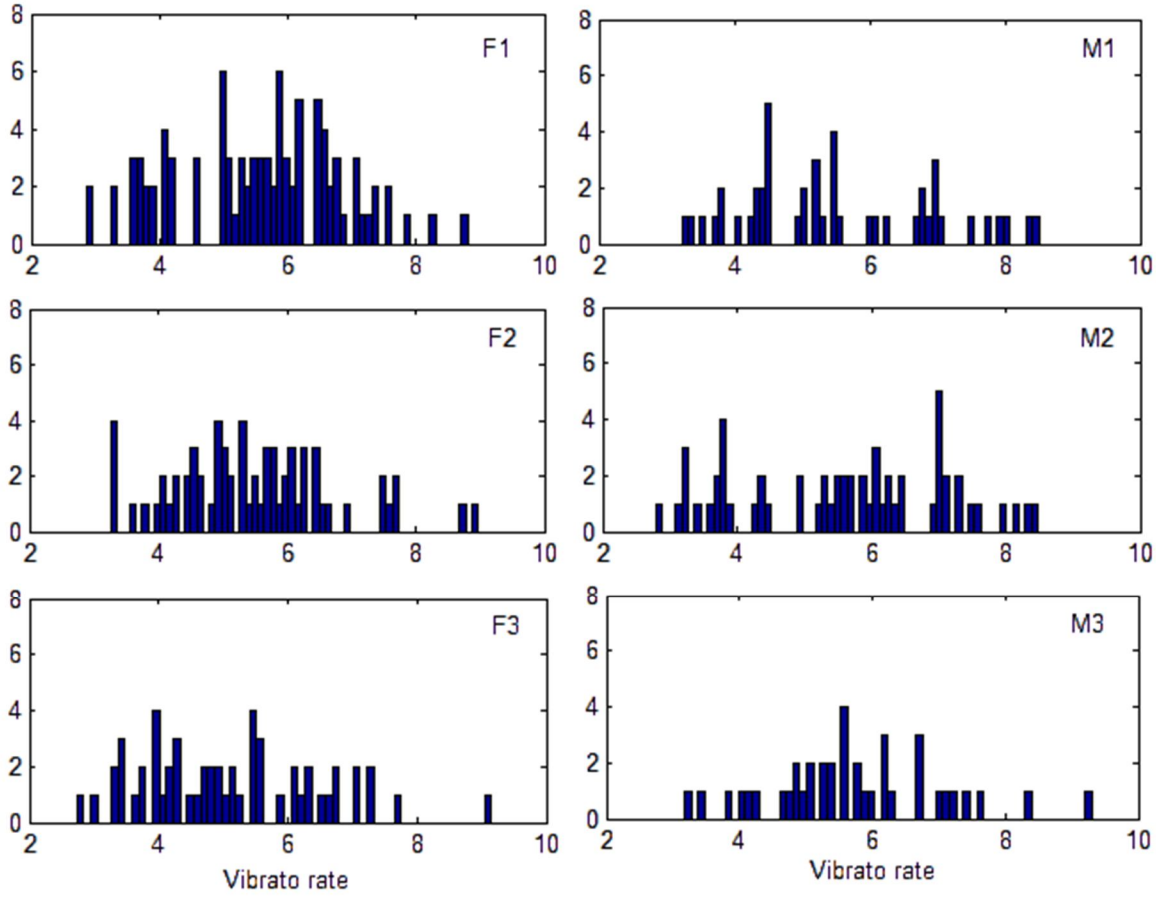


Figure 5. Histogram of vibrato rate of different singers

Analysis of Vibrato Extent

Vibrato extent values are measured for the different cappella singing voice from the database. The average extent ranges from 0.6 to 2 semitones for singers. Extent uncertainties are affected by rate uncertainties which, in turn, depend on F_o estimation, windowing, and spectral sampling [14]. The analysis of vibrato extent of six singers (three female and three male) is given in Fig. 6.

Each singer reveals their own characteristics range of vibrato extent. Singer F1 extent mostly occurs between 20 to 50 cents, whereas singer F3 covers a wide range of extent. Similar analysis is seen for male singers. The average vibrato extent values of singers are 50 cent, 52 cent, 69 cent for female singers F1, F2 and F3 respectively whereas 71 cent, 78 cent, 78 cent for male singers M1, M2 and M3 respectively. The average vibrato extent of male singers is higher than the female singers. The maximum recorded vibrato extent is found to be 159 cent for a male singer from the database of singers used for experimentation.

Combined Analysis of Vibrato Rate and Extent

Interesting observations are revealed when a singer is analyzed considering both rate and extent. Fig. 7 shows the plots of extent vs rate of all singers.

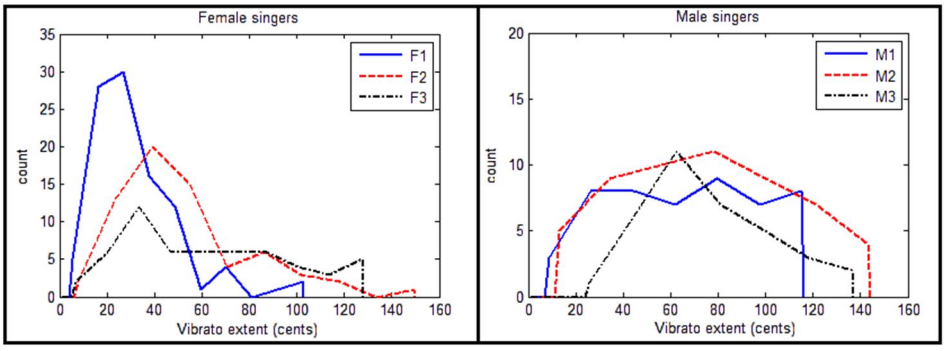


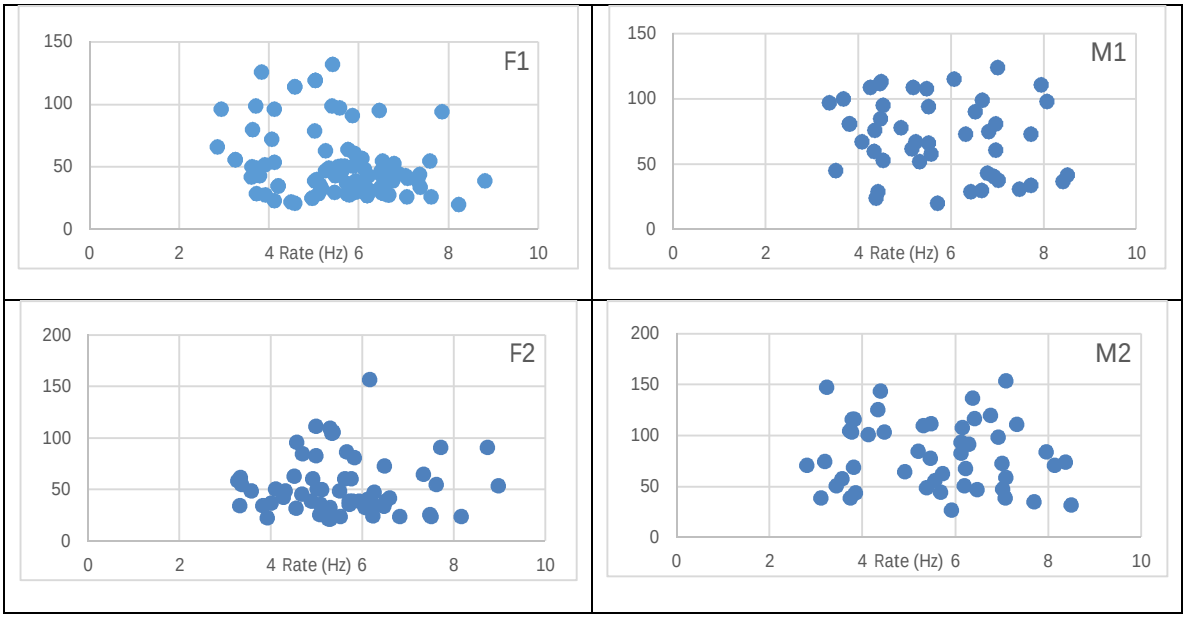
Figure 6. Vibrato extent of different singers

The observations for Singer F1 reveals that most of its vibrato information lies in a restricted range for both rate as well as extent. Similar is the observation for singers F2 and F4, whereas singers F3 and F5 are restricted in rate but have wider range of extent. As far as male singers are concerned, it is observed from Fig. 7 that singer M2 has broad range for both rate as well as extent as compared to other male singers. Thus, the plots of extent vs rate informs about the overall vibrato characteristics of individual singer.

Exploring Vibrato for Singer Identity

Every singer develops a vibrato characteristic that personalizes his/her singing style. The expressiveness of the acoustic feature has a direct impact on the Singer ID performance [8]. Fig. 7 gives a clear indication that the vibrato rate and extent patterns are different for different singers. In this research work, we explored the actual differences of these vibrato parameters among the singers for singer identity.

Table 1 shows the confusion matrix of the singer identification for 11 singers (5- Female singers and 6- Male singers) using cappella samples. Confusion matrix gives visualization of results obtained, wherein the diagonal elements represent correctly identified samples. The rows indicate the actual singers considered for identification whereas the columns indicate the hypotheses. The number of samples differ for each singer based on the cappella songs available in the database. Within the limited range of vibrato rate and extent, the maximum accuracy of singer identification achieved is almost 63%. This clearly indicates that singers develop their own vibrato naturally and subconsciously which then remains constant within the singer. Further, a clear distinction between male and female singers is achieved by incorporating a threshold on average pitch obtained from the pitch contour.



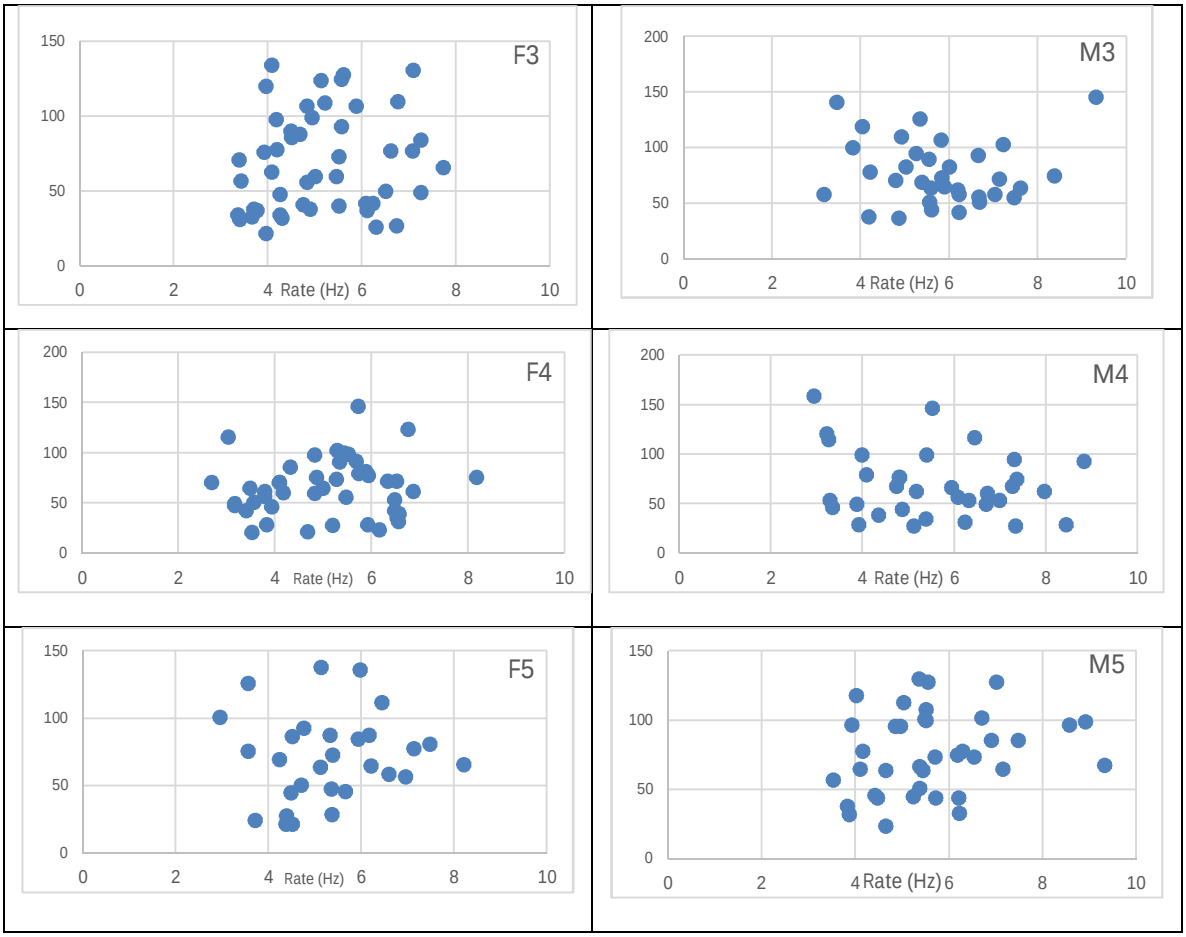


Figure 7. Vibrato rate vs Vibrato extent of different singers

Table 1.Confusion matrix for singer identification

Actual Singer Index	Hypothesized Singer Index											Accuracy (%)
	Female singers					Male singers						
	F1	F2	F3	F4	F5	M1	M2	M3	M4	M5	M6	
F1	**87**	40	19	26	7	1			1			48.06
F2	28	**54**	17	20	4							43.90
F3	10	15	**50**	13	2	2	1		1	1		52.63
F4	18	16	12	**39**	4		2	1	1			41.93
F5	8	8	5	4	**28**						1	52.83
M1	2	1	3		1	**39**	7	3	10	5	3	52.70
M2	1	1	2			12	**45**	4	8	5	8	52.52
M3						7	12	**36**	4	2	3	56.25
M4		2		3		9	13	7	**35**	2	3	47.29
M5	1		1			4	6	3	3	**34**	2	62.96
M6						8	13	5	9	2	**36**	49.31

Conclusion

This paper has described a fully automatic technique of extracting vibrato from regular cappella section of singing voice by identifying sustained pitch pieces from the song. The extracted vibrato feature i.e. rate and extent are utilized for comparatively studying the vibrato feature of different singers and explore singer identity. The analysis of vibrato parameters reveal the details of vibrato pattern of each singer and the results obtained confirm the individuality of singer. This confirms the fact that vibrato parameters remain constant for a singer and can be used for identifying singer. The singer identification obtained so far has maximum accuracy of 62%. This accuracy can be improved further (as a future scope) by combining

vibrato feature with other acoustic features of singing voice.

References

[1] Alex Loscos, "Spectral Processing of the Singing Voice", PhD Thesis, Universitat Pompeu Fabra, 2007.
[2] Wei-Ho Tsai, Hsin-Chieh Lee, "Automatic Evaluation of Karaoke Singing Based on Pitch, Volume, and Rhythm Features", IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 4, May 2012, pp. 1233-1243.
[3] Brown, W. S., Rothman, H.B., & Sapienza, C. M., "Perceptual and acoustic study of professionally trained versus untrained voices", Journal of Voice, 14 (3), 2000, pp. 301-309.
[4] Takeshi Saitou and Masataka Goto, "Acoustic and Perceptual Effects of Vocal training in Amateur Male Singing", Proc. of International Speech Communication Association, 2009, pp. 832-835.
[5] Lise Regnier, "Localization, Characterization and Recognition of Singing Voices", PhD Thesis, Universite Pierre et Marie Curie- Paris VI, March 2012.
[6] Edward Półrolniczak and Michał Kramarczyk, "Analysis of the Signal of Singing Using the Vibrato Parameter in the Context of Choir Singers", Journal of Electronic Science and Technology, vol. 11, no. 4, December 2013, pp. 417-423.
[7] Naoto Migita, Masanori Morise and Takanobu Nishiura, "A Study of Vibrato Features to Control Singing Voices", Proceedings of 20th International Congress on Acoustics, 2010, pp. 1-4.
[8] Tin Lay Nwe and Haizhou Li, "Exploring Vibrato-Motivated Acoustic Features for Singer Identification", IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no.2, Feb.2007, pp. 519-530.
[9] Justin Salamon Geoffroy Peeters, Axel R¨obel, "Statistical Characterisation of Melodic Pitch Contours and its Application for Melody Extraction", Proc. of ISMIR 2012, pp. 187-192.
[10] Peter Desain, Henkjan Honing, Rinus Aarts and Renee Timmers, "Rhythmic aspects of vibrato", Proceedings of Rhythm Perception and Production Workshop, vol.34, 1999, pp. 203-216.
[11] Tomoyasu Nakano Masataka Goto Yuzuru Hiraga, "An Automatic Singing Skill Evaluation Method for Unknown Melodies Using Pitch Interval Accuracy and Vibrato Features", Proc. of International Speech Communication Association, 2006, pp. 1706-1709.
[12] S. W. Lee, Minghui Dong and Paul Yaozhu Chan, "Analysis for Vibrato with Arbitrary Shape and its Applications to Music", Proc. of APSIPA ASC 2011, 2011.
[13] David M. Howard, Helena Daffern, Jude Brereton, "Quantitative voice quality analyses of a soprano singing early music in three different performance styles", Biomedical Signal Processing and Control 7(1), 2012, pp. 58-64 .
[14] Maurı´lio N. Vieira Jose´ Eduardo de C Silva and Hani C. Yehia, "Vibrato and tremor extent spectrum: Algorithm and applications", The Journal of the Acoustical Society of America, vol. 130, issue 1, 2011, pp.EL1-EL7.