

# Fundamental Frequency Estimation of Carnatic Music Songs Based on the Principle of Mutation

Rajeswari Sridhar , Karthiga S and Geetha T V

Department of Computer Science and Engineering  
Anna University, Chennai, India

## Abstract

Fundamental frequency estimation is very essential in Carnatic music signal processing as it is the basic component that needs to be used to determine the melody string of the signal after estimating the other frequency components. In this work a new algorithm to estimate the fundamental frequency of Carnatic music songs and film songs based on Carnatic music is proposed and implemented. The algorithm is based on the biological mutation theory which is implemented using the characteristics of Carnatic music where the concept of neutral mutations is adopted. Hence, the principle used is that, the signal characteristics do not change if it is mutated with another signal having the same frequency components. For determination of the fundamental frequency the three features namely, MFCC, spectral flux, and centroid of the original are estimated. The mutating signal is derived in a similar manner musicians adjust their singing frequency range for a particular song. The pre-recorded 'S', 'P', 'S' is used for mutating the input signal at three positions namely, beginning, middle and end. Then the same set of features namely MFCC, spectral flux, and centroid are also extracted for the mutated signal. Then by comparing the features of the original signal with the mutated signal, the signal which matches closely with the features of the original signal in all the three positions is identified and the frequency corresponding to the lower 'S' of the signal which is used for mutating is identified as the fundamental frequency of the input signal. This algorithm was evaluated using the measures of Harmonic Error, Absolute difference between mean pitches and Absolute difference in standard deviation and it was observed that the proposed algorithm yielded a better result than the existing algorithms for estimating fundamental frequency, for the input considered.

**Keywords:** *Fundamental frequency, Music signal processing, Carnatic music.*

## 1. Introduction

Fundamental frequency estimation of the audio signal is a classical problem in signal processing [1]. The estimation of fundamental frequency has been a research topic for many years both for speech and music signal processing. Fundamental frequency is the physical term for pitch [2]. Pitch is defined as the perceptual attribute of sound which

is the frequency of a sine wave that is matched to the target sound in a psychophysical experiment. Fundamental frequency is essential in speech signal processing for determining the speaker in Speaker Verification or Recognition systems. The estimation of fundamental frequency is essential in music signal processing in order to determine pitch pattern, range of pitch frequencies, music transcription, and music representation systems [1] [3]. In Carnatic music, estimating the fundamental frequency is essential and it is the foundation for determining the important characteristic of carnatic music – Raga [2]. In Carnatic music, the concept of fundamental frequency is quite different than that mentioned in speech and western music. In Carnatic music fundamental frequency refers to frequency of the middle octave 'S' which is synonymous to the note C in a keyboard in Western music. Therefore in our work, we have proposed an algorithm for the identification of fundamental frequency of Carnatic music songs based on the biological theory of mutation. The characteristics of Carnatic music have been adopted for the implementation of this mutation based algorithm.

This paper is organized as follows: Section 2 talks about some existing work in fundamental frequency estimation, Section 3 discusses about Carnatic music characteristics, Section 4 on the proposed mutation based algorithm, Section 5 on the Experimental setup and results analysis and finally Section 6 concludes the paper.

## 2. Existing Work

Fundamental frequency is defined as the lowest frequency at which a system vibrates freely and hence requires determination of the lowest frequency component from the input signal. Fundamental frequency is also defined as the reciprocal of the time period between the two lowest peak points of a given signal and hence it can also be determined by looking at the time domain representation of the signal to yield the successive lowest peak points. The signal based features that are used for the

determination of fundamental frequency can be classified as Time Domain features, Spectral features, Cepstral features and features that are motivated based on auditory theory. Many of the algorithms that are available for fundamental frequency estimation of speech and music are based on the estimation of frequency domain features and auditory motivated features.

In the algorithm developed by Arturo Camachho and John Harris [1] the authors have estimated the fundamental frequency of speech and music signal based on spectral comparisons. The average peak to valley distance of the frequency representation of the signal is estimated at harmonic locations. This value is computed at several regions of the input signal and the distance is estimated between successive average peak to valley value. Then fundamental frequency value is determined as the least distance of the average peak to valley values. This work was implemented as a combined work for speech and music. The time complexity of this algorithm is very high in the worst case situation since the distance measure needs to be computed between successive segments for all possible combinations in the input signal. Another algorithm was developed by Alain de Cheveigne and Hideki Kawahara [3] which is also a generalized algorithm for speech and music. It is based on the well-known auto-correlation method which is in turn based on the model of auditory processing. The steps involved include determining of auto-correlation value, then correcting the errors in the computed value by computing the difference function between the auto-correlation values, normalizing the value of the difference function by estimating the mean value, and iterating this correlation value to determine the fundamental frequency of the input signal. The time taken to correct the errors is very high as it is a generic algorithm for speech and music. Another algorithm developed by Robert C Maher and James Beauchamp [4], uses a two way mismatch procedure for estimating the fundamental frequency of music signals. This algorithm is based on estimating the quasi-harmonics value which requires computing the inverse square root of the fluctuating matrix and identifying the lowest value. In this algorithm, fundamental frequency is determined by computing this quasi-harmonic value for short-time spectra of the input signal. The same value is determined in the neighbouring spectra and then the fundamental frequency is estimated as the least value of the sample input segment considered. Many algorithms have also been developed for estimating multiple fundamental frequencies available in an input signal [5] [6]. In the algorithm developed by Chungshin Yeh, Axel Robel and Xavier Rodet [5] a quasi-harmonic model is developed to determine the components of harmonicity, spectral smoothness. After determining the components, a score value is assigned for the computed harmonicity value and

spectral smoothness and based on the score value the fundamental frequency is estimated. The algorithm developed by A.P. Kalpuri [6], is also based on harmonicity, spectral smoothness and synchronous amplitude evolution within a single source for determining the fundamental frequency. The authors have implemented an iterative approach where the fundamental frequency of the most prominent sound is computed and it is subtracted from the mixture and this process of computation and subtraction is iterated to determine the fundamental frequency of the signal. The authors have used spectral components like the spectral envelope and spectral smoothness. In another algorithm developed by Boris Doval and Xavier Rodet [7], the fundamental frequency of audio signal is estimated based on the evolution of the signal by assigning a probabilistic value to the pseudo-periodic signal. This algorithm developed a HMM based on the estimated spectral features to identify the fundamental frequency of the signal and hence it requires lot of training to determine the evolution of the signal. In another work proposed by Yoshifumi et al [8] the authors have identified peak values of amplitude in each segment of the frequency domain. After identifying the peak frequency based amplitude this value is represented as a sequence of pulses and auto-correlation function is applied to determine the pitch of a speech signal. This algorithm was applied only for speech signal.

All the algorithms that were developed were targeted for Western music and Speech in general and hence has used the signal characteristics of music and the characteristics of human speech for determination of the fundamental frequency. All these algorithms are also based on the fact that the lowest frequency available in the input signal is the fundamental frequency. In addition, these algorithms have high computational complexity. Since these algorithms are for determining the lowest frequency component, these algorithms are not suited for Carnatic music signal processing because of the characteristics of Carnatic music. All the algorithms proposed for Western music are under the assumption that the lowest frequency component in the input signal is the fundamental frequency but however this concept cannot be used for Carnatic music signal processing because in Carnatic music a singer can sing in two octaves that range from the mid-value of the lower octave till the mid-value of the higher octave including the middle octave [2]. Therefore the lowest frequency will not necessarily correspond to the middle octave and hence the lowest frequency cannot be assumed as the fundamental frequency for Carnatic music signal processing.

The algorithm proposed by A.P. Kalpuri [6], motivated us to use a spectral comparison based approach whereby we were motivated to move to biological theory of mutation to implement the spectral comparison algorithm. In the

algorithm implemented by [6] the authors have used spectral smoothness and harmonicity as features and they have used spectral comparisons between the segments of the same file. In our algorithm we determine features like spectral flux, centroid and MFCC and compare the input signal's features with that of the mutated input signal. For mutating the input signal the octave interval characteristics of Carnatic music is used.

### 3. Carnatic Music Characteristics

The algorithm for fundamental frequency has been implemented based on the characteristics of Carnatic music and hence it is required to explore some of the basic characteristics of Carnatic music. Carnatic music and Hindustani music are traditional Indian music systems and are very different from the traditional Western system of music. Carnatic music system is a just tempered system of music compared to the even tempered system of Western music. Just tempered system of music gives the singer the flexibility to start a particular song at any frequency as the fundamental frequency. An additional important difference is that in Carnatic music an octave has 22 intervals as against the 12 intervals of an octave system in Western music and Hindustani music [2]. The fundamental frequency normally refers to the frequency of the middle octave 'S' which is at a frequency of 240 Hz. This S corresponds to the C in Western music. In Carnatic music the singer normally starts at a frequency higher than 240 Hz and refers to this starting frequency as the 'S'. In addition, a Carnatic music song is sung in two octaves. The two octaves refer to the second half of the lower octave, the full of the middle octave and the first half of the higher octave. Hence, in order to span a frequency range of two octaves it is very important that the singer chooses the fundamental frequency with necessary caution. Hence, the frequency range of singing depends on the fundamental frequency 'f' and it ranges from '3f/4' to '3f' thus ranging over two octaves.

The need for fundamental frequency determination of Carnatic music arises because it is necessary to identify the Raga of Carnatic music piece. In Carnatic music, a Raga is defined as the sequential arrangement of the swaras. There are essentially seven swaras in Carnatic music called S, R, G, M, P, D, N which is synonymous to C, D, E, F, G, A, B in the Western music. The ascending order of the arrangement of swaras is called Arohanam and the descending order of the arrangement of swaras is called Avarohanam as given in [2]. The Raga can be classified into Parent Raga and Child Raga. A parent raga is one in which all the seven swaras are available in the Arohanam and Avarohanam. A parent raga is created by choosing all the seven swaras S, R, G, M, P, D, N. This results in a combination of 72 thereby resulting in 72

parent Ragas [2] [9] [10]. Therefore in order to determine the raga of a particular Carnatic music song it is mandatory to know the swaras available in the song which in turn is very much dependent on the fundamental frequency. Therefore depending on the starting frequency which is referred to as the frequency of the middle octave 'S' the other frequencies would slide depending on a ratio given in Table 1. Hence it becomes very essential to determine the fundamental frequency of the song to determine the various swara patterns.

Table1: Swara and their Ratio with the middle octave S

Swara	Ratio	Swara	Ratio
S	1	M2	27/20
R1	32/31	P	3/2
R2	16/15	D1	128/81
R3	10/9	D2	8/5
G1	32/27	D3	5/3
G2	6/5	N1	16/9
G3	5/4	N2	9/5
M1	4/3	N3	15/8

One more concept in Carnatic music is that the R, G, D, N can take three frequencies, M can take two frequencies and P can take only one and hence a singer can choose any one of these smaller differences in frequencies also as a fundamental frequency for singing. In addition, the frequency of the Carnatic music swaras is not discrete but is continuous. Therefore the frequency range between 240 Hz and 256.4 Hz is identified as R1, between 256.4 and 260 is termed as R2 and so on. Hence in Carnatic music the singer can choose any one of these smaller swaras also as fundamental frequency for a particular song thus resulting in a choice of one of the 22 available frequency components in an octave as the fundamental frequency.

To identify the fundamental frequency of the input song it is sufficient to identify the first few seconds of the input song as this segment has the Kalpana Swara. Kalpana Swara is the prefix tune that is sung before the beginning of the song as an aalapana or along with the Pallavi of the song [2]. Kalpana Swara is the ornamentation that is rendered to the song to emphasize the raga of the song. Therefore the Raga of the song is conveyed by the time the singer finishes the Pallavi of the song where Pallavi refers to the beginning two lines of a song [2]. In addition, Kalpana Swaras should be sung according to the fundamental frequency of the chosen song.[2]. Hence it is sufficient to consider the first few seconds of the song to identify the fundamental frequency of the input song. Therefore in our algorithm it was ensured that in the first 30 seconds of duration three or four samples of 5 seconds duration from the input song are taken for determining the fundamental frequency according to our mutation based algorithm.

## 4. Fundamental Frequency estimation

### 4.1 Basic idea of the algorithm – Mutation

The concept of mutation is a well known methodology used in many computer applications and in particular for signal processing applications [11] [12] [13]. Mutation is a phenomenon which is normally identified in a DNA molecule as a change in the DNA's sequence which is due to radiation, viruses or exposing a body to a different environment or surrounding [14]. The process of mutation which can influence the change of DNA sequence could result in an abnormality in the exposed cell. Some mutations are harmful and others are beneficial. In addition, we have a concept called as neutral mutation which does not have any effect be it beneficial or harmful but however just changes the DNA's sequence without affecting the overall structure of the DNA. The DNA is exposed to changes but this change is not causing any impact because the changed pattern is in such a way that it is one of the various combination of the existing DNA's sequence itself.

In the work done by Cristian et al [11], the authors have utilized the concept of mutation to perform genetic algorithm coding to design IIR filters. The authors have utilized the mutation operators like uniform mutation and non-uniform mutation that would select a gene from the available gene pool. After creating a gene pool, Principal Component Analysis is performed on the created pool set which is also based on the concept of mutation, "mutation tends to homogenize the components to avoid having few principal components and neglecting the others". Using the determined code values IIR filters were designed where the coefficients of IIR filters are determined using the proposed mutation technique by the authors. The authors claimed that the results of the IIR filters were better than the Newton based strategy.

In the work done by David Lu [12], the author has utilized the mutation strategy to decide the notes to be used for transcribing a piece of music. Here the authors create a gene pool of possible transcriptions for a particular piece of music and then use mutation theory that would assign a fitness value to determine the exact transcription against the possibilities of all the available transcriptions. The author has used the mutation theory of irradiate, nudge, lengthen, split, reclassify and assimilate to determine the transcription sequences.

In another algorithm done by Gustavo Reis et al [13], genetic algorithms were used for music transcription. This algorithm is similar to the one proposed by [12]. This is another algorithm in which the gene pool is iterated to determine the transcription sequence.

These algorithms for music transcription and IIR filter design motivated us to move towards mutation theory to check for the possibility of determining the fundamental frequency. In our algorithm we exploit the feature of neutral mutation to determine the fundamental frequency of the signal. The signal's frequency components are similar to the DNA's sequence. In the event of neutral mutation, the structure of DNA sequence is retained. Similarly, in our algorithm if the mutating signal is made to imbibe into the input signal, the mutated signal's frequency characteristics will be the same as the original input signal's frequency characteristics then the mutated signal and input signal would have the same set of frequency components. After mutating the signal if the signal characteristics are identical to the original signal then the fundamental frequency of the original signal is the same as the fundamental frequency of the mutating signal.

### 4.2 Algorithm and System Architecture

The pseudo code of the basic algorithm is given below.

```
Algorithm_Mutation_FundamentalFrequency(Input
Signal, Local Oscillator)
Feature Extraction (Signal .wav)
    Features = Extract MFCC, Spectral Flux, Spectral
centroid
Q1 = Q2 = Q3 = ∞
I = 1
While (Local Oscillator exists)
{
Mutate the original signal at beginning, middle and end
with the ith SPS' from the local oscillator database
    FundamentalFrequency = Frequency of ith 'S'
ModFeatures1 = Feature Extraction(Input signal Mutated
at beginning)
ModFeatures2 = Feature Extraction(Input Signal Mutated
at middle)
ModFeatures3 = Feature Extraction (Input Signal Mutated
at end)
Leastval1 = Compare ModFeatures1 with Features
Leastval2 = Compare ModFeatures2 with Features
Leastval3 = Compare ModFeatures3 with Features
If (Leastval1 < Q1 & Leastval2 < Q2 & Leastval3 < Q3)
    Then
        Q1 = Leastval1
        Q2 = Leastval2
        Q3 = Leastval3
        FundamentalFrequency = Frequency of the ith S
I = I+1
}
```

The proposed system architecture is shown in Figure 1.

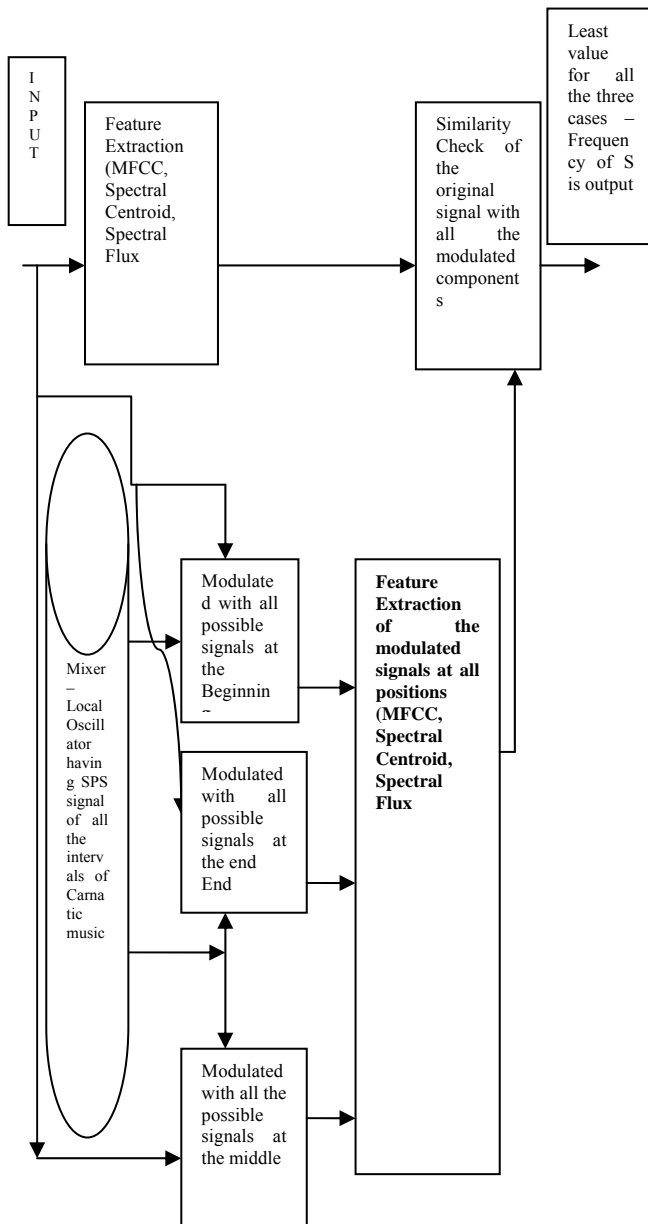


Figure 1 Architecture diagram

The various components of the architecture are the local oscillator database, the feature extractor and the comparison block. The local oscillator consists of the pre-recorded signals of 'S''P''S'' of all the 22 intervals of Carnatic music. The feature extractor is an algorithm that extracts the features like MFCC, Spectral flux and Spectral centroid. The comparison algorithm computes the distance between the extracted features of the original signal and the modified signal. The proposed algorithm is a constant time algorithm as the number of iterations this algorithm has to be evaluated is predetermined which is equal to the number of intervals of the octave of Carnatic music and is 22.

The components of the system architecture are explained in detailed below.

#### 4.2.1 Local Oscillator Database

The local oscillator database is one in which the pre-recorded 'S''P''S'' of all the intervals of Carnatic music are stored. In Carnatic music there are 22 intervals for an octave scale [2]. A Carnatic singer can start tuning their song to any one of these 22 intervals. This choice of one of the 22 intervals depends on the song that is being sung and also on the fundamental frequency of the singer. If the song has a frequency range in the lower octave than in the higher octave then this song will have a better rendering if it is being sung by a person having lower fundamental frequency and vice-versa. Hence it is obvious that singers whose fundamental frequency is not high do not choose songs which require singing in the higher octave rather than the middle octave. Also, singers whose fundamental frequency is high do not choose songs which require singing in the lower octave. Therefore in order to sing for two octaves the singer cannot choose to start singing at 400 Hz if the singer's fundamental frequency is nearly 300 Hz. Therefore, we have recorded samples of SPS starting from 240 Hz till the next octave 480 Hz. This signal is called as the mutating signal which will be imbibed into the original signal for computing the features. The process of mutating is done to the original signal in three positions. The three positions that are chosen are middle, beginning and end. The necessity of three positions arises due to the fact that the characteristics of 'S''P''S'' can occur at the beginning, end or in the middle. The three mutated signals are given individually to the feature extraction to compute the features. These 3 set of features are later compared with the features of the original signal.

#### 4.2.2 Feature Extraction

There are two feature extraction blocks in the figure. Both the blocks essentially extract the same set of features. The first feature extraction block extracts the features from the input signal. The other feature extraction block extracts the features from the mutated signal. The signal is mutated at the beginning, end and middle of the input signal, with the first SPS from the local oscillator database as explained in the algorithm explained above, thereby generating three modulated signals. Then we extract a set of features from all the modulated signals. Then the features of the mutated signal are compared with the features of the original signal. The following features are extracted.

- Mel Frequency Cepstral Coefficients (MFCC)

MFCC are based on discrete cosine transform (DCT) [15]. These coefficients are defined as the log power of the amplitude after modifying the given spectrum to a cepstrum. The mel spaced filter banks are modeled based on the perception of hearing and hence the filter banks are linear till 1000 Hz and logarithmic for the frequencies above 1000 Hz. The mel frequency cepstral coefficients are defined as:

$$C_n = \sqrt{\frac{2}{k}} \sum_{k=1}^k (\log S_k \times \cos(n(k - \frac{1}{2})\frac{\pi}{k})), n = 1, \dots, L$$

where K is the number of the subbands and L is the desired length of the cepstrum. The value of L is very small compared to K for the dimension reduction purpose and that of k is also less than K which are the filter bank energy after passing through the kth triangular band-pass filter. The input signal is converted to a frequency scale using FFT and then converted to frequency scale. After determining the mel-frequency scale, frequency bands are decided using the Mel-frequency scale and S is given by

$$S_k, 0 \leq k < K$$

- Spectral Centroid

Spectral centroid is defined as the median of the spectrum [15]. In order to determine spectral centroid, we divide the frequency band (i.e 0 to Fs/2, where F, is the sampling frequency in Hz) into a fixed number of subbands and compute the centroid for each subband using the power spectrum of the music signal.

$$C_m = \frac{\int_{f_m}^{f_{m+1}} f w_m(f) p^\gamma(f) df}{\int_{f_m}^{f_{m+1}} w_m(f) p^\gamma(f) df}$$

where P (f) is the power spectrum and  $\gamma$  is a constant controlling the dynamic range of the power spectrum. By setting  $\gamma < 1$ , the dynamic range of the power spectrum can be reduced.

- Spectral flux

Spectral flux is a measure of how quickly the power spectrum of a signal changes [15]. It is calculated by comparing the power spectrum of one frame against the power spectrum of its previous frame. More precisely, it is usually calculated as the 2-norm (also known as the Euclidean distance) between the two normalized spectra. The spectral flux that is calculated in this manner is not dependent upon overall power or on phase considerations (since only the magnitudes are compared).

$$Flux = || X_i | - | X_{i+1} ||$$

After estimating the features the input signal is now ready for comparison with the modified signal for determining the fundamental frequency of the input signal

#### 4.2.3. Comparison Check

The features are estimated and the algorithm as explained in the pseudo-code above is executed. Initially the set of features MFCC, Spectral flux and Centroid are determined from the input signal and the mutated signal. The Euclidean distance value between the original signal and that of the mutated signal is determined first with MFCC feature. If the algorithm is not able to determine the fundamental frequency of the input signal since more than one ‘S’P’S’ has the least distance value or the distance computation is not the least for all the three positions, then the spectral flux and centroid are used as features for determining the Euclidean distance between the original signal and the mutated signal to identify the fundamental frequency of the input signal.

### 5. Experimental Set up and Result Analysis

The input signal is sampled at 44.1 KHz. In our case, we estimated the fundamental frequency of nearly 100 songs and validated against the typical fundamental frequency of the singers singing range and also validated the computed result with that determined by musicologists. Tamil film music songs and Classical Carnatic music songs sung by singers like Balamuralikrishna, Ilayaraja, M.S. Subbalakshmi and Nithyasree were chosen for the purpose. Two or three samples of nearly 5 seconds from each song were chosen. The input signal was made to go through the feature extractor to extract the features like MFCC, Spectral flux and Spectral centroid. The input signal was then mutated with the local oscillator database as explained in the algorithm at the beginning, middle and end. The mutation is done at three positions because of the fact that the input signal considered is of very small duration. S, P, S can appear anywhere within the 5 seconds duration signal and hence when we consider only one position of the input signal for mutation, the frequencies S, P, S, of the mutating signal may not coexist with that of the input signal. After mutating the signal, we determine the features and run a Euclidean distance based comparison algorithm to determine the distance between the features of the original signal and the mutated signals at the three positions for all values of S, P, S from the local oscillator database. The results are tabulated and the various fundamental frequency values observed for the singers for different songs are plotted.

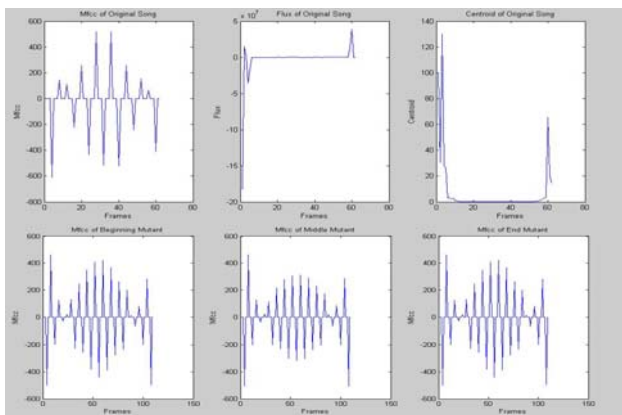


Figure 2 The input signals various features and the mutated signal's MFCC values for the SPS that has matched for Balamuralikrishna

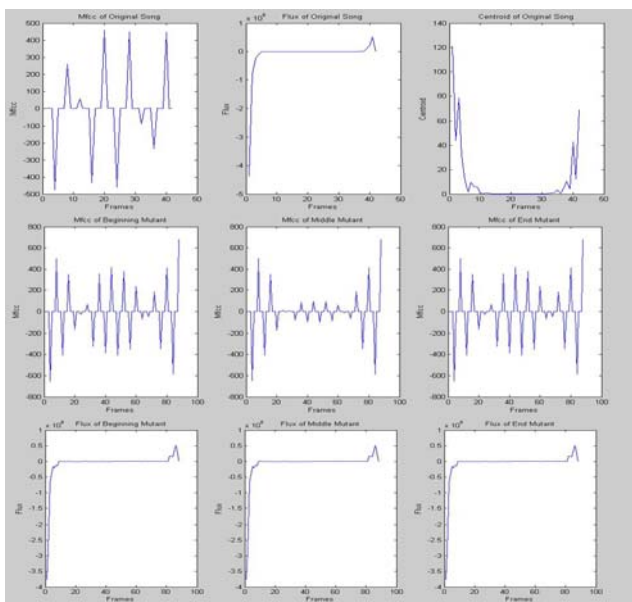


Figure 3 The input signals various features and the mutated signal's MFCC values and flux for the SPS that has matched for Ilayaraja

The Figure 2 shows the features extracted from the input signal and the corresponding mutated signals for the MFCC feature for the singer Balamuralikrishna. In this figure when MFCC coefficients were determined for the mutated signal and the original signal, the distance measure gave a least value and hence determination of fundamental frequency was done using this measure. In Figure 3, the MFCC coefficients for Ilayaraja's input were calculated and based on which a conclusion could not be reached regarding the fundamental frequency. Therefore, we took another feature, spectral flux to determine the fundamental frequency.

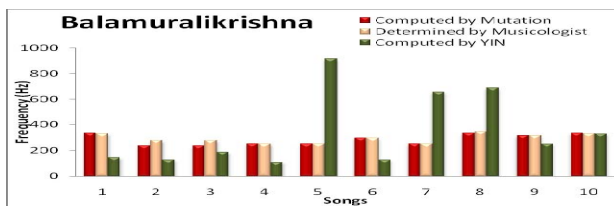


Figure 4 Fundamental Frequency Comparison of Dr. M. Balamuralikrishna

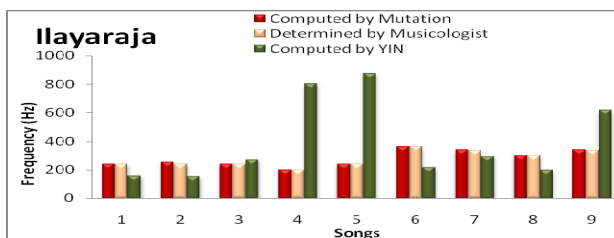


Figure 5 Fundamental Frequency Comparison of Dr. Ilayaraja

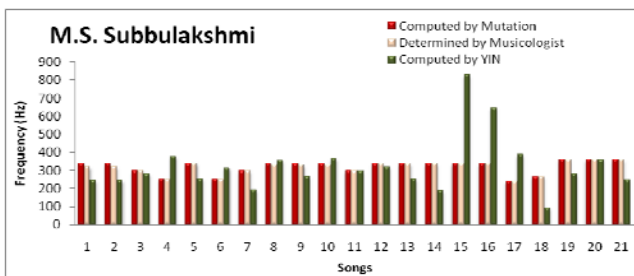


Figure 6 Fundamental Frequency Comparison of Dr. M. S. Subbulakshmi

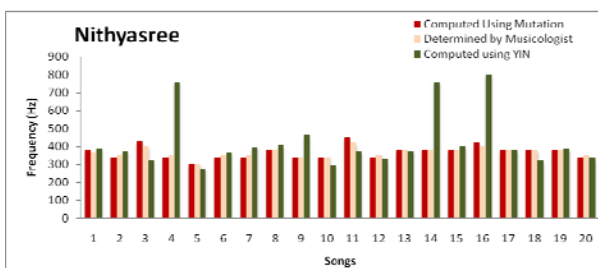


Figure 7 Fundamental Frequency Comparison of Ms. Nithyasree Mahadevan

The fundamental frequencies of different singers were determined pertaining to the different songs sung by them. The results we got were also in the range between 320 and 400 Hz. The results that were computed with the algorithm were given to Musicologists to identify the fundamental frequency of the individual songs. The same set of songs was also tested using the YIN algorithm as given in [3]. The various results for the different songs as determined by the musicologist, computed by the mutation algorithm and that of the YIN algorithm is plotted in Figure 4 to Figure 7. As can be seen from the figure, there was a

difference of 5 to 10 Hz at most between the one that is identified by the musicologists and the one that is determined by the algorithm. The normal singing frequency range of Nithyasree Mahadevan is nearly 400 Hz as suggested by Musicologists. The same observations were made for Dr. Ilayaraja, Dr. M. Balamuralikrishna and Dr. M. S. Subbulakshmi whose normal fundamental frequency of singing is 240 Hz, 320 Hz, and 330 Hz respectively. The sudden peaks that were observed in their fundamental frequencies are due to the fact that the sample chosen started at the higher octave rather than at the middle octave. Another sample in the same song yielded a different fundamental frequency. This ambiguity was later resolved by computing the distance measures on multiple samples and the sample value which gave the least difference in distance for a given 'S''P''S' is identified as the fundamental frequency. However when the results were compared between the output of the YIN algorithm and that of the mutation algorithm it was observed that YIN algorithm's output always determined the lowest frequency component available in the input or the harmonics value of the lowest frequency of the input song. This argument was the one which was put forward to justify a separate algorithm for Carnatic music fundamental frequency determination. The output of the YIN algorithm actually corresponds to the lower octave D, N as against the middle octave S.

## 5.1 Algorithm Evaluation

The Mutation based algorithm and the YIN algorithm were implemented to determine the fundamental frequency of four singers and the algorithm was evaluated for the following parameters based on the evaluation suggested in [16] by Bojan Kotnik et al. The authors have used the already proposed parameters by Joseph Martino et al [17] and Ying et al for parameters like gross error high, gross error low, voiced errors, unvoiced errors, average mean difference in pitch and average difference in standard deviation. All these parameters estimated the percentage in difference between the actual frequency and the computed frequency by considering the speech signal as a voiced and unvoiced signals. Another algorithm proposed by [18] estimated precision and recall and also F-measure for evaluating the fundamental frequency. All these measure gave an estimate of identifying the correct frequency against a wrong frequency as fundamental. This motivated us to introduce a new evaluation parameter based on the observed results. We term this parameter as harmonic frequency estimation error. Some of the algorithms that are already available for speech and western music mostly gave the harmonic of the lowest frequency and hence we used this as one more parameter for evaluation. In addition we also used parameters like average mean difference in

pitch and average difference in standard deviation. The parameters are discussed below

### 5.1.1 Harmonic frequency Estimation Error (HE)

In Western music or Speech in general the lowest frequency component is termed as the fundamental frequency or pitch. However in Carnatic music, the lowest frequency component is not the fundamental frequency as already explained. When comparison of the algorithms of YIN and mutation based were made, it was observed that YIN determined the Harmonic of the lowest frequency in more number of situations than the mutation based algorithm. This is because of the voiced and unvoiced components being present in the input music piece. When the fundamental frequency is available in the unvoiced component segment this frequency is skipped and the algorithm identified the harmonic of the fundamental frequency [16]. The Harmonic frequency estimation error is defined as the ratio of harmonic of the fundamental estimated as against the determination of the fundamental frequency. The determination of the harmonic error is important for Carnatic music signal processing since the determination of fundamental frequency is important for identification of swara pattern thereby leading to Raga identification. If the harmonic is identified as the fundamental frequency then it would result in the wrong swara pattern. In addition, the fundamental frequency indicates the singing range of the singer. For example, if the harmonic frequency is 500 Hz as against the fundamental 250 Hz then it indicates the singing range from 250 Hz to 1500 Hz as against 125 to 750 Hz. Therefore determination of this error is essential in determining the singing range of the singer and help in correct Raga identification.

The harmonic error is estimated for the four singers based on the fundamental frequency identified which is listed in the Figures 4 to Figures 7. The results are tabulated and the performance chart is given in Figure 8. As can be seen in the figure the mutation based algorithm and the one determined by musicologists have a low error rate in estimating the harmonic frequency as against the actual fundamental frequency when compared to YIN.

The problem with the singer Ilayaraja is the singing range is at a low fundamental frequency between 200 Hz to 240 Hz, hence the mutation algorithm also identified the harmonic of the fundamental frequency. As can be observed in the YIN algorithm where the lowest frequency is termed as the fundamental frequency, the probability of the harmonic frequency being identified as the fundamental frequency as against the actual fundamental frequency is high.



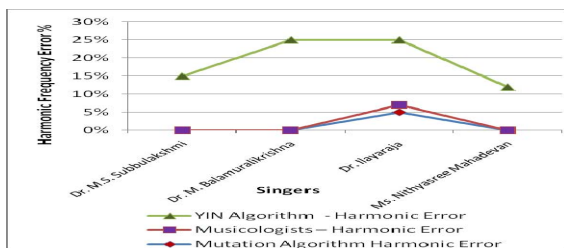


Figure 8 Harmonic Frequency Estimation Error

### 5.1.2 Absolute difference between mean values (ABDM)

The absolute difference (in Hz) between the mean values (ABDM) of the reference fundamental frequency which is the normal singing range of the singers and the actual estimated fundamental frequency is estimated as given in [16] as

$$ABDM[Hz] = \text{abs} \{ \text{MeanRefPitch}[Hz] - \text{MeanEstPitch}[Hz] \}.$$

The average fundamental frequency as estimated by the mutation algorithm, YIN algorithm and musicologists were determined and the reference pitch was chosen as 400Hz, 320 Hz, 300Hz, 250 Hz for Nithyasree, M.S. Subbulakshmi, Balamuralikrishna and Ilayaraja respectively. The reference pitch is chosen by observing their normal range of singing and the absolute difference between the mean values is estimated and is given in Figure 9.

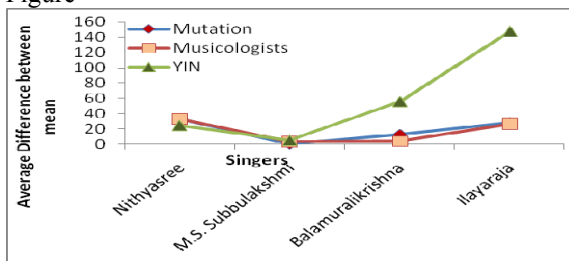


Figure 9 Absolute difference between mean values

It was observed that for singers Nithyasree, M.S. Subbulakshmi the estimations by YIN algorithm and mutation algorithm were comparable while for singer Balamuralikrishna the YIN algorithm gave a higher difference between the computed value and the reference value.

### 5.1.3 Absolute difference between standard deviations (AbsStdDiff)

The absolute difference (in Hz) between the standard deviations of reference fundamental frequency and the actual estimated fundamental frequency is computed as given in [16] by

$$AbsStdDiff[Hz] = \text{abs} \{ \text{StdRef}[Hz] - \text{StdEst}[Hz] \}$$

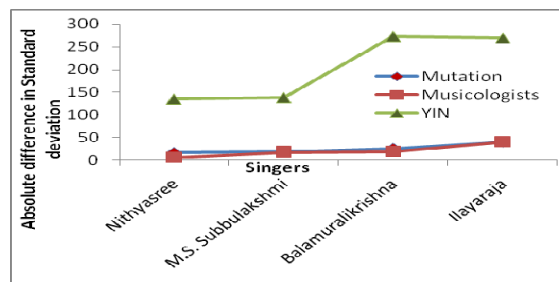


Figure 10 Absolute differences between standard deviation

The mentioned mean values and standard deviations are computed on whole reference and estimated F0 data respectively. The graph showing the absolute difference between standard deviation is plotted in Figure 10. It is observed that the YIN algorithm deviated to a greater extent from the other two algorithms since the YIN algorithm computed the harmonic of the lowest frequency in most of the situations for all the singers. The mutation algorithm as well as the one determined by musicologist was almost the same for all the singers.

## 6. Conclusion and Future Work

In this work a new system for identification of Fundamental frequency of Carnatic music songs and film music songs based on Carnatic music is implemented and tested. This algorithm was compared with the existing algorithm of YIN. It was observed that the YIN algorithm always determined the lowest frequency component of the input signal as the fundamental frequency which is not the case always in Carnatic music processing. This work of identification of Fundamental frequency is an essential module for Signal processing applications of speech and music. It is even more mandatory for Carnatic music processing because of the just tempered behavior of the system. In the just tempered behavior system the frequencies belonging to one Raga pattern is based on the fundamental frequency of the song depending on the time of rendering the song. Hence, after estimating the fundamental frequency this module can be integrated with our already determined algorithm for Raga identification and Singer identification. In our earlier system, the fundamental frequency was assumed to be 240 Hz and used to determine the raga of a given song. Now this system, with integration of the fundamental frequency module can be computed on the fly and hence the raga can be determined in a dynamic manner. Our earlier Singer identification system used a set of coefficients called as CICC [9] which were based on fundamental frequency. Now these coefficients would be identified with the identification of fundamental frequency making it a robust Singer identification system.

## References

- [1] Arturo Camacho and John G Harris, "A Sawtooth waveform inspired pitch estimator for speech and music", Journal of the Acoustical society of America, 2008
- [2] Prof. P Sambamurthy, "South Indian Music" Vol 1 – 6, The Indian Music Publishing house, India.
- [3] Alain de Cheveigne and Hideki Kawahara, "YIN a fundamental frequency estimator for speech and music", Journal of Acoustical society of America, 2002
- [4] Robert C Maher and James Beauchamp, "Fundamental frequency of music signals using a two way mismatch procedure", Journal of Acoustical society of America, 1994
- [5] Chungshin Yeh, Axel Robel and Xavier Rodet, "Multiple fundamental frequency estimation of polyphonic music signals", ICASSP 2005
- [6] A.P. Kalpuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness", IEEE Transactions on Speech and Audio Processing, 2003
- [7] Boris DOVAL and Xavier RODET, "Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and HMM", ICASSP 1993
- [8] Yoshifumi Haraa, Mitsuo Matsumotob and Kazunori Miyoshia, "Method for estimating pitch independently from power spectrum envelope for speech and musical signal", J. Temporal Des. Arch. Environ. 9(1), December, 2009
- [9] Rajeswari Sridhar and T.V. Geetha, "Raga Identification of Carnatic music for Carnatic music Information Retrieval", International Journal of Recent trends in Engineering, 2009
- [10] Rajeswari Sridhar and T.V. Geetha, "Music information retrieval of Carnatic songs based on Carnatic music singer identification, IEEE International Conference on Computer and Electrical Engineering, 2008
- [11] Cristian Munteanu and Vasile Lazerescu, "Improving mutation capabilities in a real-coded genetic algorithm", First European Workshop on Evolutionary Computation in Image Analysis and Signal Processing, EvoIASP '99
- [12] David Lu, "Automatic Music Transcription using Genetic algorithms and Electronic synthesis", Thesis of David Lu.
- [13] Gustavo Reis, Nuno Fonseca, Francisco Fernandez de Vega, Anibal Ferreira, "Hybrid Genetic algorithm based on gene fragment competition for polyphonic music transcription", Springer LNCS, December 2008
- [14] Howard Ochman, "Neutral Mutations and Neural substitution in Bacterial Genomes", Society for Molecular Biology and Evolution, 2003.
- [15] Rabiner and Juang, "Fundamentals of Speech Recognition", Prentice Hall Signal Processing Series, 1993
- [16] Bojan Kotnik, Harald Höge<sup>2</sup>, Zdravko Kacic, " Evaluation of Pitch detection algorithms in adverse conditions", Proc. 3<sup>rd</sup> International Conference on Speech Prosody, Dresden, Germany, pp. 149-152, 2006.
- [17] Martino, J., Yves, L., " An Efficient F0 Determination Algorithm Based on the Implicit Calculation of the Autocorrelation of the Temporal Excitation Signal", Proc. EUROSPEECH'99, Budapest, Hungary.
- [18] Ying, G., Jamieson, H., Mitchell, C., "A Probabilistic Approach to AMDF Pitch Detection", Proc. ICSLP 1996, Philadelphia, PA.
- [19] Mert Bay Andreas F. Ehmann J. Stephen Downie, "EVALUATION OF MULTIPLE-F0 ESTIMATION AND

TRACKING SYSTEMS", 10<sup>th</sup> International Society for Music Information Retrieval 2009, Kobe, Japan

**Rajeswari Sridhar** is a Sr. Lecturer in the Department of Computer Science and Engineering at Anna University, Chennai, India. She is currently doing her Ph.D in the area of Carnatic music signal processing. She has nearly 10 publications to her credit and is interested in the areas of Signal Processing, Theoretical Computer Science, Language Technologies

**S. Karthiga** is a PG student of the Department of Computer Science and Engineering at Anna University Chennai. Her areas of interest include, Data Structures, Algorithms and Information retrieval

**Dr. T. V. Geetha** is a Professor in the Department of Computer Science and Engineering at Anna University, Chennai, India. She has more than 20 years of teaching experience and has produced 10 Ph.D students so far and is guiding 11 students. She has more than 180 publications to her credit and is interested in the area of Tamil Computing. She is currently working on Language technologies and managing sponsored and funded projects.