Objective Assessment of Ornamentation in Singing

Submitted in partial fulfilment of the requirements
of the degree of

Master of Technology
(*Communication & Signal Processing*)

by

Chitralekha Gupta
(08307R05)

under the guidance of

Prof. Preeti Rao

Department of Electrical Engineering
Indian Institute of Technology Bombay
2011

Dedication

I dedicate this thesis to my family. Without their patience, understanding, support and most of all love, the completion of this work would not have been possible.

Approval Sheet

This dissertation entitled **Objective assessment of ornamentation in singing** by **Chitralekha Gupta** (Roll no. 08307R05) is approved for the degree of Master of Technology in Electrical Engineering.

Prof. Preeti Rao ________________________ (Supervisor)

Prof. B. K. Dey ________________________ (Examiner)

Dr. Samudravijaya K. ________________________ (Examiner)

Prof. R. K. Joshi ________________________ (Chairperson)

June 20th, 2011

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Chitralekha Gupta
08307R05

June 20th, 2011

Acknowledgment

I express my sincere gratitude towards Prof. Preeti Rao for her constant guidance and support throughout the course of this project. The regular discussions with her on every aspect of the research work helped me refine my approach towards the problem and motivated me to give my best. Working with her in the field of audio signal processing was a very pleasant and learning research experience and my interest in the subject has considerably grown.

I would like to thank Vishweshwara Rao for his valuable suggestions and help during various stages of my project. The technical inputs as a result of numerous discussions with him have been critical in this project work. I would also like to thank Sachin Pant for his help in making the perceptual test user interfaces and scoring demonstration interface. I thank all the members of the Digital Audio Processing lab, Department of Electrical Engineering, IIT Bombay for providing a friendly and enjoyable working environment.

I acknowledge the role of all the singers and the judges for their help in preparation of database and validation of the system. I would also like to thank the music teacher Mrs. Ranade, the classical musicians Pt. Nayan Ghosh and Dhanashree Pandit Rai and the research scholars from ITC SRA, Kolkata for their valuable inputs during the course of the project.

Finally I thank my family for their love and moral support throughout my course duration.

I thank all the people who have contributed ideas, concepts and corrections to be incorporated in my project.

Chitralekha Gupta

Abstract

Important aspects of judging a singing performance include musical accuracy and voice quality. In the context of Indian classical music, not only is the correct sequence of notes important to musical accuracy but also the nature of pitch transitions between notes. These transitions are essentially related to *alankars* (ornaments) that embellish the inherent beauty of the genre. Thus a higher level of singing skill involves achieving the necessary expressiveness via correct rendering of ornamentation, and this ability can serve to distinguish a well trained singer from an amateur. In this work, we explore objective methods to assess the quality of ornamentation rendered by a singer with reference model or ideal singer of the same song. Methods are proposed for the perceptually relevant comparison of complex pitch movements, and validated with respect to subjective ratings by human experts. Such an objective assessment system can be a valuable feedback tool in the training of amateur singers and also be used for competitive singing platforms that aim to identify good singers from the masses.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

DFT	Discrete Fourier Transform
ER	Energy Ratio
RER	Relative Energy Ratio
ED	Euclidean Distance
DTW	Dynamic Time Warping
FDOscAmp	Frequency Domain Oscillation Amplitude
FDOscRate	Frequency Domain Oscillation Rate
TDOscAmp	Time Domain Oscillation Amplitude
TDOscRate	Time Domain Oscillation Rate
ZCR	Zero Crossing Rate
CART	Classification and Regression Tree

Chapter 1. Introduction

1.1. Motivation

Evaluation of singing quality could be done by note matching, and expression scoring. While learning how to sing, the first lessons from the guru (teacher) involve training to be in *sur*, that basically means to hit the right notes (key matching). Even the primary criteria while judging a singer is his/her capability of rendering the notes correctly. In the context of Indian Classical Music, not only the sequence of correct notes but also the nature of transitions between notes that are essentially related to *alankars* (ornaments), is considered to be important as it embellishes the inherent beauty of the genre. So, the next grade of singing involves how well can the singer improvise on the rendered notes that makes the song more expressive and pleasing to hear. The degree of perfection in rendering such expressions gives an important cue to judge the singers above the level of key matching that will distinguish between an amateur from a well trained singer. So incorporating expression scores in the singing evaluation systems for Indian music in general is expected to increase its performance in terms of its accuracy with respect to perceptual judgment. Such a system will be useful in singing competition platforms that involve screening out better singers from large masses. Also such an evaluation system could be used as a feedback tool for training amateur singers.

1.2. Problem definition

The aim of this work is to formulate a method for objective evaluation of singing quality based on closeness of various types of expression rendition of a singer to that of the reference or the ideal singer. An equally important problem is to evaluate singing quality without a reference audio available. However this problem is not considered in the present work. Several methods to evaluate a specific ornament extracted from sung phrases have been explored and subjective judgment has been used to validate the methods.

1.3. Organization

A brief outline of the report is described as follows. In Chapter 2, the literature in this area of research has been reviewed and the important musical concepts used in this work are summarized. A brief system overview and the database used for the experiments as well as the subjective tests have been described in Chapter 3. The experimental assessment of the three different ornaments along with the validation results and discussion has been described in Chapter 4, 5 and 6. The conclusions of this work and possible future work have been discussed in Chapter 7.

Chapter 2. Literature Review

Pitch represents the perceived fundamental frequency of a sound. The pitch of singing voice is determined by the rate of vibration of the vocal cords. Vocal cord vibration results in puffs of air that result in pressure variations, which reach our ears as sound. Singing sounds consist of integral multiples of pitch or fundamental frequency. The fundamental frequency range of the singing voice is much larger than that of speech. For males, the fundamental frequency can vary from 70 to 500 Hz and for females, from 150 to 700 Hz [**1**].

Pitch perception is closely correlated with the periodicity or fundamental frequency. Human pitch perception is approximately logarithmic with respect to fundamental frequency, i.e. constant pitch changes in music refers to a constant ratio of fundamental frequencies. The perceived distance between the pitches 220 Hz and 440 Hz is the same as the perceived distance between the pitches 440 Hz and 880 Hz. Human ear is sensitive to ratios of fundamental frequencies (pitches), not to absolute pitch. Motivated by this logarithmic perception, music theorists sometimes represent pitches using a numerical scale based on the logarithm of fundamental frequency. A cent is a logarithmic unit of measure used for musical intervals. Twelve-tone equal temperament divides an octave into 12 semitones of 100 cents each. 1200 cents are equal to one octave — a frequency ratio of 2:1 — and an equally tempered semitone (the interval between two adjacent piano keys) is equal to 100 cents.

$$Pitch_{cents} = 1200 \log_2\left(\frac{Pitch_{Hz}}{440}\right) \tag{2.1}$$

where 440 Hz tone serves as the audio frequency reference. It is the internationally recognized standard for musical pitch that serves as reference for the calibration of pianos, violins, and other musical instruments.

Melody is the temporal sequence of discrete notes (pitches). But for singing voice, this pitch contour is not just made up of discrete horizontal lines corresponding to distinct steady notes but with a continuously evolving curve that have macro and micro-tonal pitch movements and ornamentations. These comprise the expressive elements of music.

2.1. Western Musical Expressive Elements

In western music, ornaments can be considered to be all extra-note events. The musical expressive elements in Western singing mainly include *vibrato*, *glissando*, and *tremolo.* For instance, attaining good vibrato requires prolonged training by professional singing coaches. Moreover, while prominent singers' vibrato is easily distinguishable, this is not the case for beginners. Assessment of vibrato in Western singing has been extensively explored [**2**][**3**][**4**] and has been reviewed in this work.

Vibrato is a musical effect produced in singing and in playing of musical instruments, serving to enrich the musical sound. It is defined as a deliberate, periodic fluctuation of pitch that can be parameterized by its rate (the number of vibrations per second) and extent (the amplitude of vibration from an average pitch on the vibrato section). It is used to add expression and vocal-like qualities to instrumental music. Empirically, vibrato commonly has a periodicity of 5-8 Hz and extent range is 30 – 150 cents [**3**] (Figure 2.1).

Though it is a western musical expressive element, vibrato is widely used in Indian popular music. One such example is shown in Figure 2.2**.**

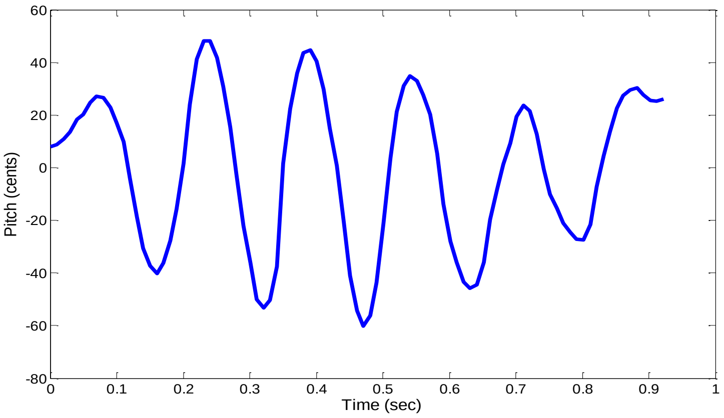


Figure 2.1 Vibrato

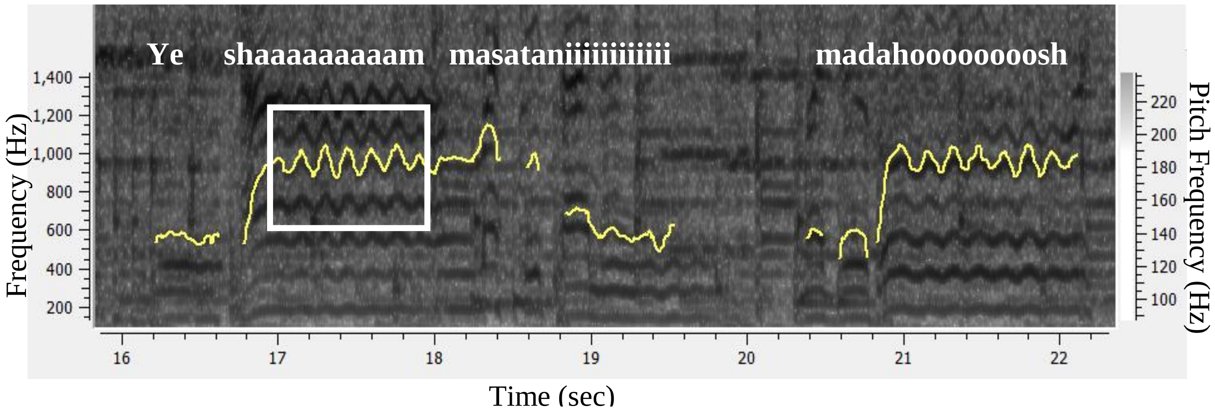


Figure 2.2 Spectrogram and vocal pitch contour of a phrase from the song *Ye Shaam* from the movie *Kati Patang.*The ornament *vibrato* is marked by a box. Lyrics of the phrase are written at the top.

2.2. Indian Musical Expressive Elements

In the context of Indian Classical Music, the manner in which notes are linked and embellished is as important as the correct use of steady notes. All the extempore variations that a performer creates during a performance within the *raga* and *tala* limits could be termed as *alankar* (ornament), because these variations embellish and enhance the beauty of the *raga*, the *tala* and the composition. Style, emotions, *gharana* characteristics, *raga* characteristics and even the personal characteristics might be embedded in these ornaments. Of the many ornaments, the ones that appear in the transcriptions are *Meend, Andolan, Khatka, Murki, Gamak, Zamzama* [**5**] [**6**]. Out of these, the possibility of scoring the ornaments resembling *meend* and *gamak* has been explored in this study.

Following is a brief description of the above mentioned ornaments as given in [**5**] and [**6**]. These concepts have been extracted from a musicological source but there are differences in opinion between different musicologists regarding the nature of these ornaments and their exact definition. The pitch plots and spectrograms of the ornament clips shown here are entirely based on the way these ornaments are interpreted in this work. All the excerpts of audio shown in the spectrograms and plots are taken from Indian film songs that are rich in ornaments. The lyrics of the phrase shown in the spectrograms are written at the top. Some of the examples are excerpts from an interview with the famous Hindustani classical vocalist Dhanashree Pandit Rai. The ornaments marked in them are annotated by her.

i. *Meend*

A *meend* is a glide from one note to another. Proper rendition involves: accuracy of starting and ending notes, speed of the *meend*, and accent on intermediate notes (Figure 2.3). The duration of this ornament is approximately 1 sec. There are three types of *meends*:

1. straight forward, smooth, unidirectional – ascending or descending
2. with a slight pause on one or more intermediate notes
3. undulating *meend* – up and down, wave-like movement

Also a *meend* may originate from a note that is only fleetingly touched – i.e. a grace note or a *kan-swar* – that enhances the beauty of the succeeding notes of the *meend*. An incorrect *kan-swar* with *meend* would bring in shades of some other raga into the performance.

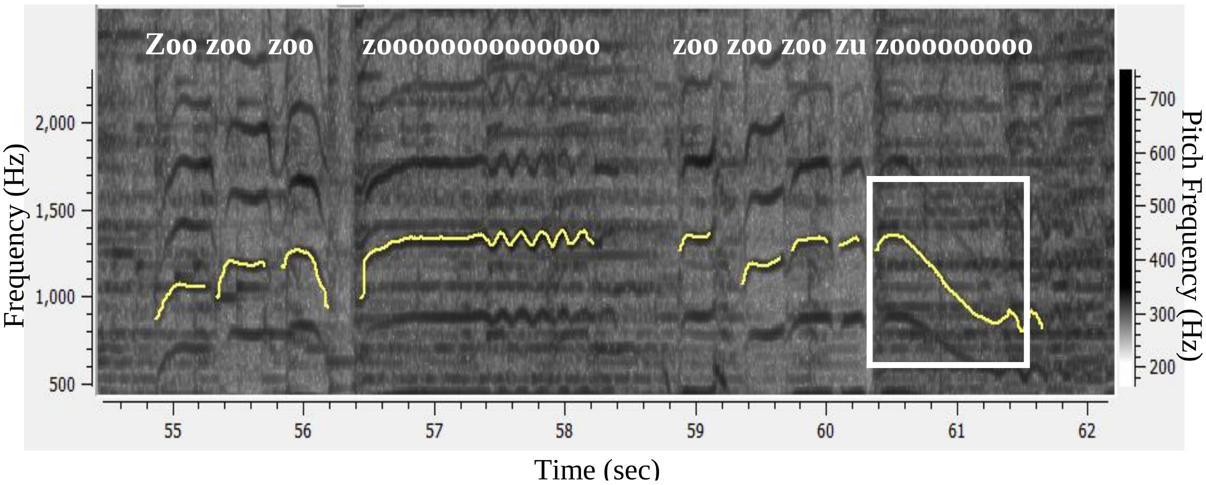


Figure 2.3 Spectrogram and vocal pitch contour of a phrase from the song *Kaisi paheli* from the movie *Parineeta.*The ornament *meend* is marked by a box.

ii. Gamak

A *gamak* can be defined as a fast *meend* (spanning 2-3 notes normally) delivered with deliberate force and vigor and repeated in an oscillatory manner. This is a relatively difficult ornament to render (Figure 2.4). The duration of this ornament is less than 1 sec. It is observed that this ornament is particularly a giveaway when differentiating good/trained singers from amateur singers.

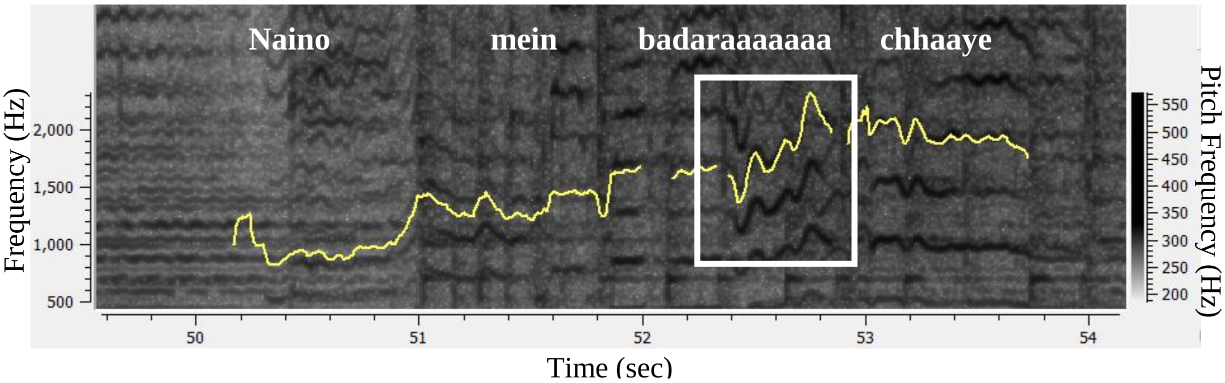


Figure 2.4 Spectrogram and vocal pitch contour of a phrase from the song *Naino Mein* from the movie *Mera Saaya.*The ornament *gamak* is marked by a box.

iii. Khatka

When a knot or cluster of notes is sung or played very fast and with gusto to decorate or embellish another note, it is called a *khatka* or *gitkari* (Figure 2.5). This is a short duration ornament of less than 0.5 seconds.

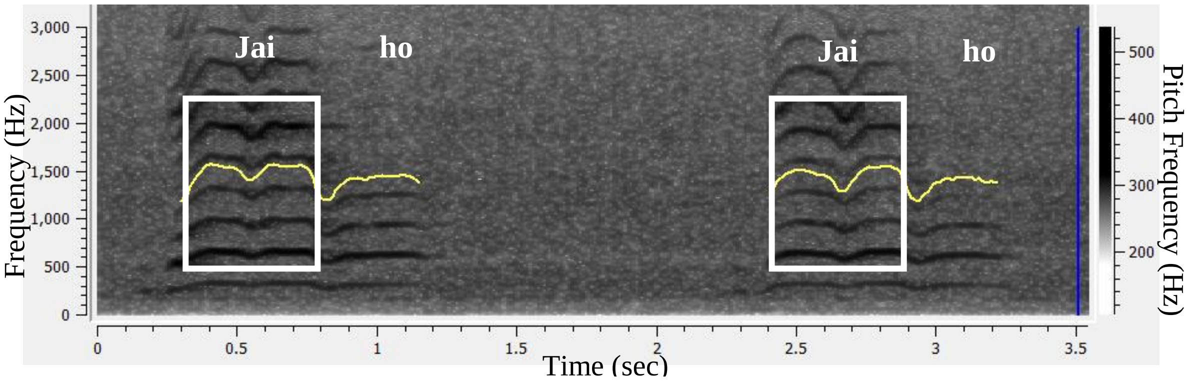


Figure 2.5 Spectrogram and vocal pitch contour of a phrase from the song *Jai Ho* from the movie *Slumdog Millionaire*, sung by Dhanashree Pandit Rai.The ornament *khatka* is marked by a box.

iv. Murki

A *murki* is a fast delicate ornament involving two or more notes (Figure 2.6).

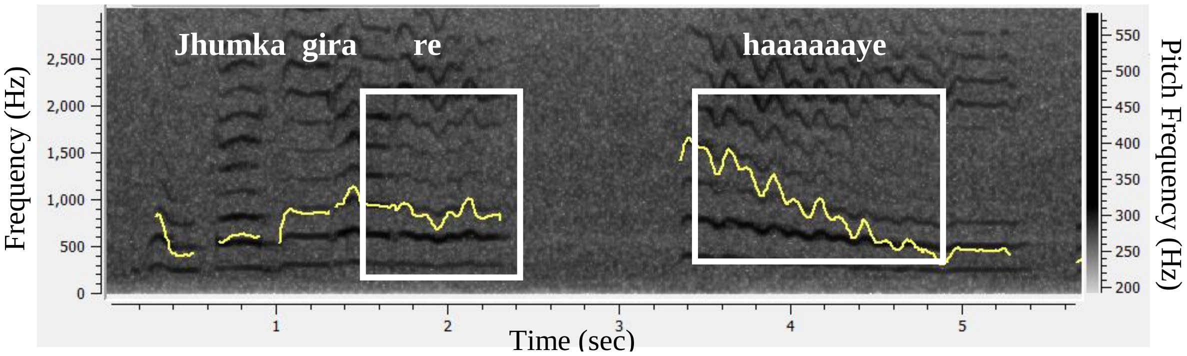


Figure 2.6 Spectrogram and vocal pitch contour of a phrase from the song *Jhumka gira* from the movie *Mera Saaya*, sung by Dhanashree Pandit Rai. The ornament *murki* (“re”) and *zamzama* (“haaaaye”) are marked by a box.

v. Andolan

The *andolan* alankar is a gentle swing or oscillation that starts from a fixed note and touches the periphery of an adjacent note (Figure 2.7). The duration of the ornament shown is 4 seconds approximately.

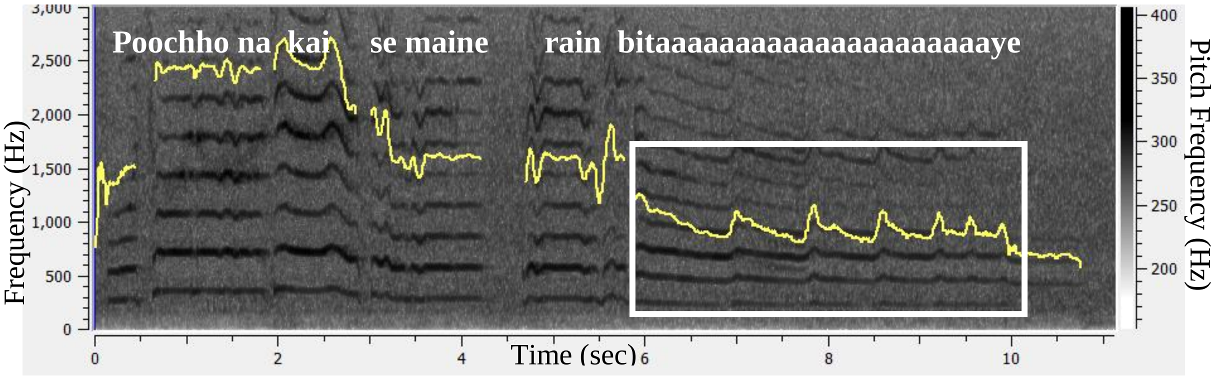


Figure 2.7 Spectrogram and vocal pitch contour of a phrase from the song *Poochho na kaise* from the movie *Meri soorat teri aankhen*, sung by Dhanashree Pandit Rai.The ornament *andolan* is marked by a box.

vi. Zamzama

Zamzama is a cluster of notes rendered in progressive combinations and permutations (Figure 2.6). The duration of this ornament is approximately 1.5 sec.

In Indian classical singing, quantitative assessment of progress of music learners based on improvisation, form and content is a topic under research [**7**]. There has been some objective analysis of the ornament *meend*, resembling a glide. Its proper rendition involves the accuracy of starting and ending notes, speed, and accent on intermediate notes [**5**][**6**]. Perceptual tests to differentiate between synthesized singing of vowel /a/ with a pitch movement of falling and rising intonation (concave, convex , linear) between two steady pitch states, 150 and 170 Hz, using a second degree equation, revealed that the different types of transitory movements are cognitively separable [**8**]. A methodology for automatic extraction of *meend* from the performances in Hindustani vocal music described in [**9**] also uses the fit to a second degree equation to detect the *meend*. Also automatic classification of *meend* attempted in [**10**] gives some important observations like frequency of occurrence of descending *meends* is maximum, and then that of rise-fall *meends* (*meend* with *kan swar*). The *meends* with intermediate touch notes are relatively less frequent. The duration of *meend* is generally between 300 - 500 ms.

In this work, three ornaments have been considered viz., vibrato, glide and oscillations-on-glide. The assessment is with respect to the “model” or ideal rendition of the same song. Considering the relatively easy availability of singers for popular Indian film music, we use Hindustani classical music based movie songs for testing our methods. Though vibrato is predominantly an ornament used in Western style of singing, it is common in popular Indian film music. The method given in [**2**], for assessment of the quality of vibrato has been investigated. The previous work reported on glide has been to model it computationally. In this work, computational modelling has been used to assess the degree of perceived closeness between a given rendition and a reference rendition taken to be that of the original playback singer of the song. The possibility of modelling more complex ornaments like oscillations-on-glide has also been explored.

Chapter 3. Methodology

This chapter gives an overview of the research methodology including the audio datasets used to develop, and eventually validate, the objective scoring algorithms proposed in this thesis. Apart from recording a number of test singers, it was necessary to obtain subjective judgements of the singing quality from expert listeners such as musicians or music teachers. Reference pitch contours of selected ornaments are extracted from original audio recordings of selected songs of famous playback singers.

3.1. Pitch Extraction from Reference Audio Signals

From the polyphonic reference audio files as well as the monophonic test audio files, the pitch contour is extracted using a semi-automatic polyphonic pitch detection tool [**11**] that computes pitch every 10 ms interval throughout the audio segment. The parameters that were accessible in the tool and frequently used for manual adjustments in the pitch contour for reliable extraction are as follows:

i. *Pitch range low (Hz):* Lower limit on F0 search range.

ii. *Pitch range high (Hz):* Upper limit on F0 search range. Typical vocal range for speech is known to be 80 -350 Hz. But for music the range is more. For Female singers a reasonable range is 125 to 600 Hz although some classical singers even pitch at 850 Hz. For male singers the pitch range is usually from 70 to 500 Hz.

iii. *Frame length (sec):* depends on two factors: For singers with higher pitch range, use a smaller frame length (20 ms) and for singers with lower pitch range use longer windows (40 ms). A good trade-off is 30 ms. Frame length also depends on the rate of pitch modulations in the audio excerpt. For more rapid modulations, as is prevalent in taans in Hindustani music, use shorter frame lengths (20 ms).

iv. *rho - Octave bias :* This biases the final pitch values towards lower or higher octaves. Higher values (0.15 - 0.25) bias the pitch values towards lower octaves to track male voice pitch and lower values (0.05 - 0.1) bias final pitch values towards higher octaves to track female voice pitch.

v. *Pitch jump cost:* This parameter controls the penalty applied for pitch jumps. For audio segments where pitch is not changing rapidly a value of 0.1 is recommended. For audio segments with rapid pitch variations, larger values (0.2 - 0.3) may be used as this will reduce the penalty on large pitch jumps.

vi. *Voicing threshold:* This is for voicing detection i.e. detection of the presence of singing voice. This will mark those pitch values above a given threshold as voiced. It is seen that setting this value between -18 to -30 yields acceptable results for all types of music.

The pitch contour obtained was validated in two ways:

i. By observing the similarity of the harmonic trajectories visible in the spectrogram with the obtained pitch contour.

ii. By resynthesizing the pitch contour by constant energy method, available in the interface, and validating the contour perceptually. This method of resynthesis generates a complex tone with three equal amplitude harmonics. This method is also used to generate the pitch resynthesized version for the subjective tests as described in the Section 3.4.2.

The details about the datasets and annotation are provided next.

3.2. Reference Data

The dataset consisting of polyphonic audio clips from popular Hindi film songs rich in ornament, were obtained as the reference dataset. The ornament clips (300 ms – 1sec) were isolated from the songs for use in the objective analysis. Short phrases (1 - 4 sec duration) that include these ornament clips along with the neighbouring context were used for subjective assessment. The ornament clips along with some immediate context makes it perceptually more understandable.

3.3. Test Data

The reference songs were sung and recorded by 5 to 7 test singers. The test singers were either trained or amateur singers who were expected to differ mainly in their expression abilities. The method of ‘sing along’ with the reference (played at a low volume on one of the headphones) at the time of recording was used to maintain the time alignment between the reference and test songs.

3.4. Subjective Assessment

The original recording by the playback singer is treated as the model, with reference to which singers of various skill levels are to be rated. The subjective assessment of the test singers for performed by a set of 3 - 4 judges who were asked either to rank or to categorize (into good, medium or bad classes) the individual ornament clips of the test singers based on their closeness to the reference ornament clip.

3.4.1. Kendall’s Coefficient

Kendall's W (also known as Kendall's coefficient of concordance) is a non-parametric statistic that is used for assessing agreement among judges [**12**]. Suppose that object *i* is given the rank $r_{i,j}$ by judge number *j*, where there are in total *n* objects and *m* judges. Then the total rank given to object *i* is

$$R_i = \sum_{j=1}^{m} r_{i,j} \tag{3.1}$$

and the mean value of these total ranks is

$$\overline{R} = \frac{1}{2} m(n+1) \tag{3.2}$$

The sum of squared deviations, *S*, is defined as

$$S = \sum_{i=1}^{n} (R_i - \overline{R})^2 \tag{3.3}$$

and then Kendall's *W* is defined as

$$W = \frac{12S}{m^2(n^3 - n)} \tag{3.4}$$

Kendall's W ranges from 0 (no agreement) to 1 (complete agreement).

3.4.2. Subjective Relevance of pitch contour

Since all the objective evaluation methods are based on the pitch contour, a comparison of the subjective evaluation ranks for two versions of the same ornament clips - the original full audio and the pitch re-synthesized with a neutral tone, can reveal how perceptual judgment is influenced by factors other than the pitch variation.

Table 3.1 shows inter – judge rank correlation (Kendall Coefficient W) for a vibrato and a glide segment. Correlation between the two versions' ranks for each of the judges ranged from 0.65 to 1 with an average of 0.84 for the vibrato clip and 0.67 to 0.85 with an average of 0.76 for the glide clip. This high correlation between the ratings of the original voice and resynthesized pitch indicate that the pitch variation is indeed the major component in subjective assessment of ornaments. We thus choose to restrict our objective measurement to capturing differences in pitch contours in various ways.

Table 3.1 Agreement of subjective ranks for the two versions of ornament test clips (original and pitch re-synthesized)

Ornament Instance	**No. of Test Singers**	**No. of Judges**	**Inter-judges' rank agreement (W) for**		**Avg. correlation between original and pitch re-syn. judges' ranks (W)**
			Original	**Pitch re-synthesized**	
Vibrato	5	4	0.64	0.50	0.84
Glide	5	4	0.86	0.76	0.76

3.5. Methodology

The methodology for our study on objective assessment of ornaments is explained in this section. Figure 3.1 gives the implementation of the system for objective measurement of a test singer.

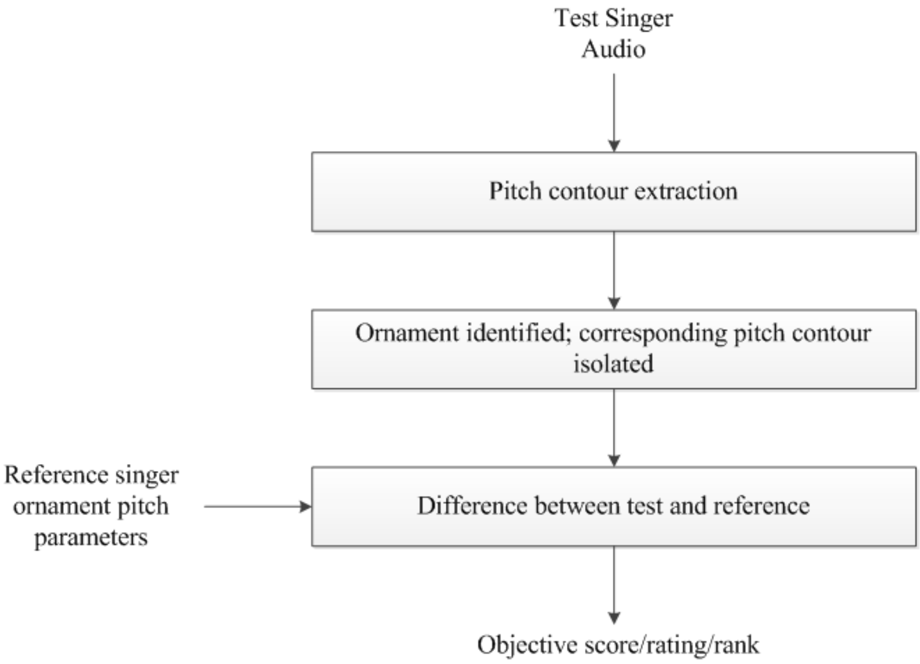

Test Singer
Audio
Pitch contour extraction
Ornament identified; corresponding pitch contour isolated
Reference singer ornament pitch parameters
Difference between test and reference
Objective score/rating/rank

Figure 3.1 System Overview

i. From the reference polyphonic and test monophonic audio files, first the pitch contour is extracted using the PolyPDA tool, as explained in Section 3.1.

ii. The ornament is first identified in the reference and marked using the software PRAAT, and the corresponding ornament segment pitch is isolated from both the reference and the test singer files for objective analysis. Also slightly larger segment around the ornament is clipped from the audio file for the subjective tests so as to have the context.

iii. Model parameters are computed from the reference ornament pitch.

iv. Subjective ranks/ratings of the ornaments for each test token compared with the corresponding reference token are obtained from the judges. Those ornament tokens that obtain a high inter-judge agreement (Kendall's W>0.5) are retained for use in the validation of objective measures.

v. The ranks/ratings are computed on the retained tokens using the objective measures for the test ornament instance in comparison to the reference or model singer ornament model parameters.

vi. The subjective and objective judgments are then compared by computing a correlation measure between them.

The procedure for modelling and rating of each of the three ornaments – vibrato, glide and oscillations-on-glide, have been examined, implemented and validated in the subsequent chapters. The pitch contour extracted from the audio has been used for modelling and computing the objective measures.

Chapter 4. Vibrato Assessment

The ornament vibrato is a well known ornament in the Western style of singing and it is well explored and modelled in the literature as explained in [**2**], [**3**], [**4**]. The method mentioned in [**4**] applies autocorrelation (after removal of DC) and the Fourier transform to the pitch contour of vibrato and explores several potential measures for assessing its quality like height of absolute maximum peak in the spectrum function, height of first autocorrelation peak, energy between 4.5 - 7.5 Hz as compared to energy between 1 - 10 Hz, frequency of absolute maximum peak in the spectrum function, height of first autocorrelation trough, lag of first autocorrelation peak, lag of first autocorrelation trough, number of peaks with amplitude larger than 1/4 of absolute maximum peak in the spectrum function. Out of these features, the selected features of highest importance were absolute height of highest peak above 2 Hz in the DFT of the pitch contour, and energy between 4.5 and 7.5 Hz that had a good correlation with the judges' average ratings.

In the following work, vibrato assessment has been done as a first step to understand ornament modelling and rating. Since the vibrato segments have sinusoidal looking pitch contour and is periodic, so a direct DFT of the segment (without autocorrelation) and the top two features from the literature as mentioned above have been explored to assess the quality of vibrato. This chapter consists of the database description followed by description of the objective measures, their validation and discussion.

4.1. Database

This section consists of the reference data, test singing data and the subjective rating description. The database used for assessing this ornament was rather small because the experiments done on this ornament were based on the work previously done mainly to formalize the procedure of ornament assessment.

4.1.1. Reference and Test Datasets

Two datasets, A and B, consisting of polyphonic audio clips from a popular Hindi film song rich in vibrato, was obtained as presented in Table 4.1. The pitch tracks of the ornament clips were isolated from the songs for use in the objective analysis. The ornament clips (1 - 4 sec) from Dataset A and the complete audio clips (1 min. approx.) from Dataset B were used for subjective assessment as described later in this section. The reference songs were sung and recorded by 5 - 10 test singers (Table 4.1).

Table 4.1 Vibrato database description

Datset & Song No.	Song Name	Singer	No. of ornament clips	No. of Test singers	Total no. of test tokens	Characteristics of the ornaments
A1.	Kaisi Paheli (Parineeta)	Sunidhi Chauhan	3	5	15	Duration of vibrato clips is 0.5 – 1 second
B1.	Pal Pal (Black Mail)	Kishore Kumar	10	10	100	
B2.	Dilbar Mere (Satte pe Satta)	Kishore Kumar	6	10	60	
B3.	Ye Shaam (Kati Patang)	Kishore Kumar	7	7	49	

4.1.2. Subjective Assessment

4.1.2.1. Dataset A

The subjective assessment of the test singers for Dataset A was performed by 4 judges who were asked to rank the individual ornament clips of the test singers based on their closeness to the corresponding reference ornament clip. The audio clips for the ornament vibrato were of approximately 3 - 4 seconds that comprised of a complete phrase whose last note had a vibrato. The judges were asked to focus on this last note and rank the test singers' clips relative to the reference, i.e. the test clip perceptually closest to the reference was ranked 1 and so on.

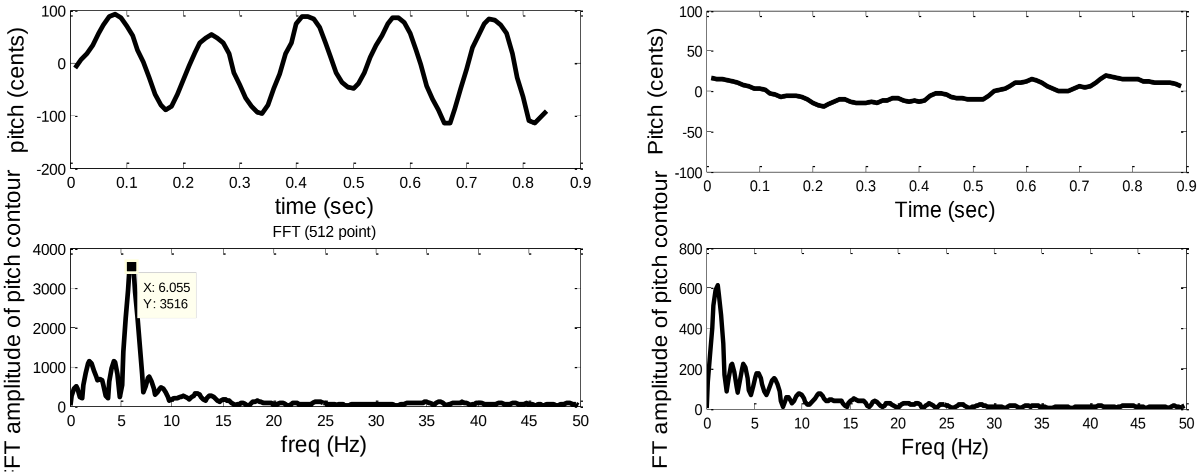
4.1.2.2. Dataset B

The subjective evaluation of the test singers for Dataset B was performed by 3 judges who were asked to categorize the test singers into one of three categories (good, medium and bad) based on an overall judgment of their ornamentation skills as compared to the reference by listening to the complete audio clip. There are 6 – 10 vibrato instances per audio clip. The inter-judge agreement was observed to range from 0.84 to 1 for the three songs' test singer sets.

4.2. Objective Measures

The pitch contour (in cents) of the entire vibrato segment (approx. 1 sec.) sampled at every 10 ms is mean subtracted versus time segment and the 512 point magnitude spectrum is processed for the objective measures. First, 9 different samples of vibrato segment from the reference singer's rendition were taken to examine their characteristics. It was observed that the frequency of peak after 2 Hz in the magnitude spectrum was consistently between 6.05 to 6.64 Hz. Also the ratio of the energy in 4.5 – 7.5 Hz to that in 1 – 10 Hz is very high, mostly > 0.7. Figure 4.1 shows the vibrato and its spectrum plot for the reference singer and a test singer ranked lowest by judges. The test singer's spectrum plot indicates the absence of vibrato.

For rating the test singers, the best two features as obtained from [**2**] and explained in the following subsection, have been used. For both the measures, a higher value is considered to be a better vibrato rendition hence ranked higher. Here, the reference vibrato parameters are not used in the objective evaluation.



(a) (b)

Figure 4.1 Pitch contour and magnitude spectrum of vibrato segment of (a) Reference singer (b) Test Singer

4.2.1. Absolute peak value in magnitude spectrum after 2 Hz

The amplitude of the vibrato could be related to the strength of the peak of the DFT between 2 Hz to 10 Hz.

$$P = \max\left(\left| Z\left(k\right) \right|_{k_{2Hz} \le k \le k_{10Hz}} \right) \tag{4.1}$$

where $Z(k)$ is the DFT of the mean-subtracted pitch trajectory $z(n)$ and k_{fHz} is the frequency bin closest to f Hz.

4.2.2. Energy Ratio

The concentration of energy in the range 4.5 Hz – 7.5 Hz (the range of rate of vibrato) could be related to the ratio of energy in the region 4.5 Hz – 7.5 Hz to that in 1 – 10 Hz. Energy ratio (*ER*) is given as

$$ER = \frac{\sum_{k=k_{4.5Hz}}^{k_{7.5Hz}} \left|Z(k)\right|^2}{\sum_{k=k_{1Hz}}^{k_{10Hz}} \left|Z(k)\right|^2} \tag{4.2}$$

where $Z(k)$ is the DFT of the mean-subtracted pitch trajectory $z(n)$ and k_{fHz} is the frequency bin closest to f Hz. ER quantifies the purity of the vibration in terms of its sinusoidal nature. It is not influenced by the frequency extent of the vibrato.

4.3. Validation Results and Discussion

A single overall subjective rank is obtained by ordering the test singers as per the *sum* of the individual judge ranks. Spearman Correlation Coefficient (ρ), a nonparametric (distribution-free) rank statistic that is a measure of correlation between subjective and objective ranks, has been used to validate the system. If the ranks are x_i, y_i, and $d_i = x_i - y_i$ is the difference between the ranks of each observation on the two variables, the Spearman rank correlation coefficient is given by [**13**]

$$\rho = 1 - \frac{6\sum {d_i}^2}{n(n^2-1)} \tag{4.3}$$

where, n is the number of ranks. ρ close to -1 is negative correlation, 0 implies no linear correlation and 1 implies maximum correlation between the two variables.

4.3.1. Dataset A

The overall subjective ranks and intra – judges' rank correlation (Kendall Coefficient *W*) as well as the objective ranks as obtained from the two objective measures and their correlation with subjective ranks (ρ) have been shown in Table 4.2 .

Table 4.2 Judges' avg. rank and objective measures 1 and 2 for three vibrato instances for 5 test singers. Measure 1: Absolute peak value in magnitude spectrum after 2 Hz, Measure 2: Energy Ratio W is Kendall's coeff. (agreement among judges), $\rho_{(ab)}$ and $\rho_{(ac)}$ are Spearman rank correlation between judges' avg. rank and measure 1 & 2 respectively

Test Singers	**Vibrato instance 1**			**Vibrato instance 2**			**Vibrato instance 3**		
	Judges' Avg. Rank (a)	**Measure 1 Rank (b)**	**Measure 2 Rank (c)**	**Judges' Avg. Rank (a)**	**Measure 1 Rank (b)**	**Measure 2 Rank (c)**	**Judges' Avg. Rank (a)**	**Measure 1 Rank (b)**	**Measure 2 Rank (c)**
Chi	2	3	2	5	3	3	2	3	2
Gau	4	5	4	4	5	5	4	5	5
Gir	5	4	1	2	4	2	2	4	4
Mbs	3	1	5	3	1	4	4	2	3
Rij	1	2	3	1	2	1	1	1	1
	W = 0.64	$\rho_{1(ab)}$ = 0.6	$\rho_{1(ac)}$ = -0.2	W = 0.64	$\rho_{2(ab)}$ = 0.3	$\rho_{2(ac)}$ = 0.7	W = 0.61	$\rho_{3(ab)}$ = 0.5	$\rho_{3(ac)}$ = 0.7

Measure 2 (Energy Ratio) shows a good correlation with subjective ratings except in vibrato instance 1 (Table 4.2). In vibrato instance 1, singers 'Chi' and 'Gir' sing at approximately the right frequency, but the amplitude of vibration in 'Chi' is more than that in 'Gir', that clearly shows up in Measure 1 (the absolute peak value after 2 Hz) while Measure 2 (energy ratio) fails to show this in this case because of the peak close to 1 Hz in DFT plot of 'Chi'. Also all the listeners have uniformly ranked 'Chi' better than 'Gir' (Figure 4.2 Vibrato instance1 and its magnitude spectrum of test singers (a) Chi and (b) Gir). This problem is fixed by tuning the frequency range of the denominator of Energy ratio measure to 2 Hz – 10 Hz so as to eliminate near DC frequencies.

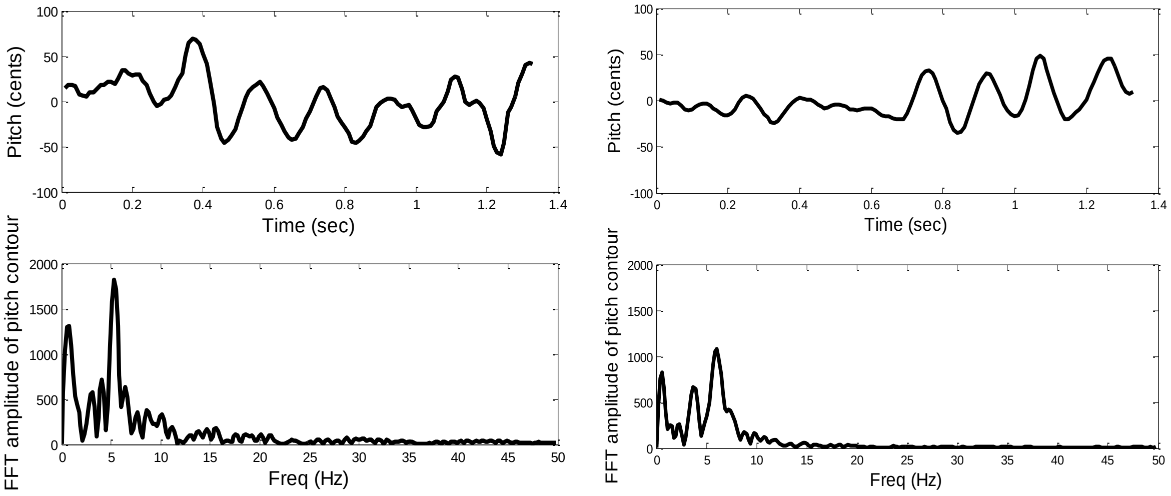
So far, the objective measures have not considered the reference ornament parameters. We now consider this. For comparing the test ornament with respect to the reference ornament, the ratio of the modified energy ratio parameter (denominator of ER set as 2 Hz – 10 Hz) of the test vibrato (ER_{Test}) to that of the reference vibrato (ER_{Ref}) is computed and ranked. Henceforth this measure is referred to as Relative ER Measure (RER) given by

$$RER = \frac{ER_{Test}}{ER_{Ref}} \tag{4.4}$$

The results using the RER measure are tabulated in Table 4.3.

Table 4.3 Judges' avg. rank and RER measure for three vibrato instances for 5 test singers. *W* is Kendall's coeff. (agreement among judges), $\rho_{(ab)}$ is Spearman rank correlation between judges' avg. rank and RER measure

Test Singers	**Vibrato instance 1**			**Vibrato instance 2**			**Vibrato instance 3**		
	Judges' Avg. Rank (a)	**RER measure**	**RER measure Rank (b)**	**Judges' Avg. Rank (a)**	**RER measure**	**RER measure Rank (b)**	**Judges' Avg. Rank (a)**	**RER measure**	**RER measure Rank (b)**
Chi	2	0.92	2	5	0.58	3	2	0.74	3
Gau	4	0.76	4	4	0.48	4	4	0.64	5
Gir	5	0.84	3	2	0.95	2	2	0.84	1
Mbs	3	0.43	5	3	0.38	5	4	0.72	4
Rij	1	0.923	1	1	0.97	1	1	0.78	2
	W = 0.64		$\rho_{1(ab)}$ = 0.6	W = 0.64		$\rho_{2(ab)}$ = 0.6	W = 0.61		$\rho_{3(ab)}$ = 0.8



(a) (b)

Figure 4.2 Vibrato instance1 and its magnitude spectrum of test singers (a) Chi and (b) Gir

One point of confusion in vibrato analysis was the evaluation of singers who sang flat against singers who sang unsteady notes, which the judges said "seems nervous and shaky, rather than vibrato - it sounds unpleasant". Figure 4.3 shows the spectrum of the pitch contour for such a singer. From the figure it can be seen that although the frequency for peak in the spectrum is 7.03 Hz, which is close to the ideal vibrato rate, a significant amount of energy lies at other frequencies as well. The energy ratio measure gives a higher rank to this singer as compared to a singer with no vibrato. To evaluate 'nervousness/shakiness", the 8th parameter mentioned in [**4**] viz. "Number of peaks with amplitude larger than 1/4 of absolute maximum peak in the spectrum function" can be explored. But there still remains the confusion whether or not to mark such cases as bad.

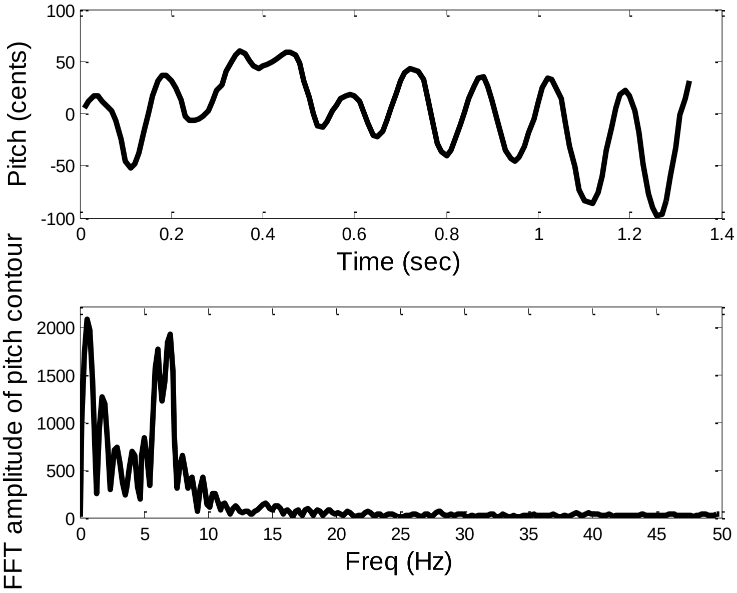


Figure 4.3 Vibrato instance1 and its magnitude spectrum of test singers Rij

4.3.2. Dataset B

The overall ornament quality evaluation of the singer as evaluated on Dataset B has good inter-judge agreement for almost all singers for all the songs in this dataset. The most frequent rating given by the judges (two out of the three judges) for a singer was taken as the subjective ground truth category for that singer.

The RER measure is used for evaluating each of the vibrato instances for the songs in Dataset B. A threshold of 0.5 was fixed on this measure to state the detection of a particular vibrato instance. For a test singer, if all the vibrato instances are detected, the singer’s overall objective rating is “good”; if the number of detections is between 40 – 100% of the total number of vibrato instances in the song, the singer’s overall objective rating is “medium”; and if the number of detections is less than 40%, the singer’s overall objective rating is “bad”. The above settings are empirical. Table 4.4 shows the singer classification confusion matrix. There are no drastic misclassifications such as good singer objectively classified as bad or vice versa. The overall correct classification is 66.7%. This clearly indicates the feasibility of objective assessment of singers based on their ornamentation skills.

Table 4.4 Singer classification confusion matrix for Dataset B

Objectively→ **Subjectively↓**	**G**	**M**	**B**
G	3	1	0
M	0	9	6
B	0	2	6

4.4. Summary

The RER measure shows a good correlation with the subjective ratings for both micro and global ratings by the subjects (Datasets A and B). Since the subjects were asked to rate the test singers compared to the reference singer, hence we have come up with a relative objective measure (RER). But since the ornament vibrato is characterized by a standard range of rate and extent, so intuitively the absolute ER measure should also give a good correlation with the subjective ratings and comparison with the reference should not be required.

Chapter 5. Glide Assessment

A glide is a pitch transition ornament that resembles the ornament *meend*. Its proper rendition involves the following: accuracy of starting and ending notes, speed, and accent on intermediate notes [**6**]. Some types of glide are shown in Figure 5.1.

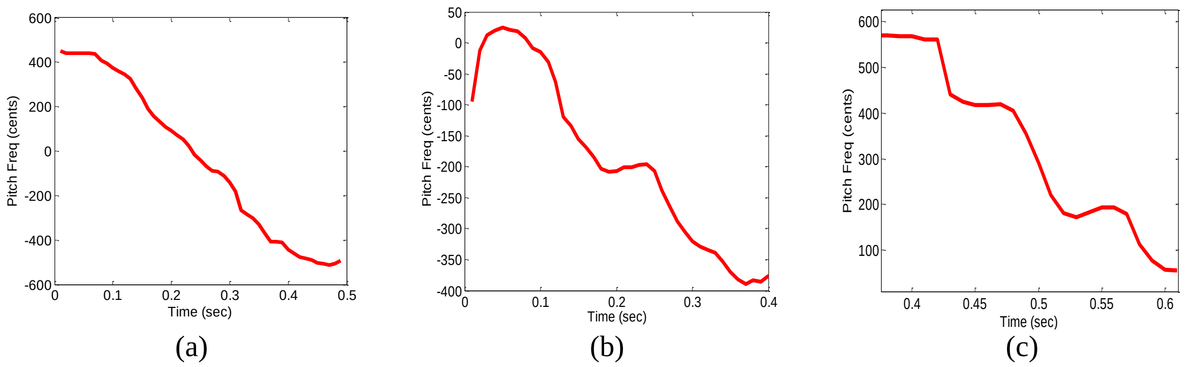


Figure 5.1 Types of *Meend* (a) simple descending (b) pause on one intermediate note (c) pause on more than one intermediate notes

This chapter explains the different objective measures used to evaluate this ornament and their validation using subjective ratings from human experts.

5.1. Database

This section consists of the reference data, test singing data and the subjective rating description.

5.1.1. Reference and Test Dataset

Two datasets, A and B, consisting of polyphonic audio clips from popular Hindi film songs rich in ornaments, were obtained as presented in Table 5.1. The pitch tracks of the ornament clips were isolated from the songs for use in the objective analysis. The ornament clips (1 - 4 sec) from Dataset A and the complete audio clips (1 min. approx.) from Dataset B were used for subjective assessment as described later in this section. The reference songs of the two datasets were sung and recorded by 5 to 9 test singers (Table 5.1).

Table 5.1 Glide database description

Data set & No.	Song Name	Singer	No. of ornament clips	No. of Test singers	Total no. of test tokens	Characteristics of the ornaments
A1.	Kaisi Paheli (Parineeta)	Sunidhi Chauhan	3	5	15	All the glides are simple descending (avg. duration is 1 sec approx.)
A2.	Nadiya Kinare (Abhimaan)	Lata Mangeshkar	4	5	20	All are descending glides with pause on one intermediate note (avg. duration is 0.5 sec approx.)
A3.	Naino Mein Badra (Mera Saaya)	Lata Mangeshkar	3	6	18	All are simple descending glides (avg. duration is 0.5 sec approx.)
A4.	Raina Beeti Jaye (Amar Prem)	Lata Mangeshkar	4	7	28	First and fourth instances are simple descending glides, second and third instances are complex ornaments (resembling other ornaments like *murki*)
B1.	Ao Huzoor (Kismat)	Asha Bhonsle	4	9	36	All are simple descending glides
B2.	Do Lafzon (The Great Gambler)	Asha Bhonsle	4	8	32	All are simple descending glides

5.1.2. Subjective Assessment

The original recording by the playback singer is treated as ideal, with reference to which singers of various skill levels are to be rated.

5.1.2.1. Dataset A

The subjective assessment of the test singers for Dataset A was performed by 3 judges who were asked to rank the individual ornament clips of the test singers based on their closeness to the reference ornament clip. The audio clips for the ornament glide comprised of the start and end steady notes with the glide in between them. The judges were asked to rank order the test singers' clips based on perceived similarity with the corresponding reference clip.

5.1.2.2. Dataset B

The subjective evaluation of the test singers for Dataset B was performed by 4 judges who were asked to categorize the test singers into one of three categories (good, medium and bad) based on an overall judgment of their ornamentation skills as compared to the reference by listening to the complete audio clip. The inter-judge agreement was 1.0 for both the songs' test singer sets.

5.2. Objective Measures

For evaluation of glides, three methods to compare the test singing pitch contour with the corresponding reference glide contour are explored: (i) point to point error calculation using Euclidean distance, (ii) Dynamic Time Warping (DTW) and (iii) polynomial curve fit based matching.

5.2.1. Point to point error calculation

Point to point error calculation using Euclidean distance is the simplest approach (Figure 5.2(a)). Euclidean distance (ED) between two points p_i and q_i is calculated as

$$d(p,q)=\sqrt{\sum_{i=1}^{n}(p_i-q_i)^2} \tag{5.1}$$

But the major drawback of this method is that it might penalize a singer for perceptually unimportant factors because a singer may not have sung 'exactly' in the same way as the reference and yet could be perceived to be close and scored high by the listeners.

5.2.2. Dynamic Time Warping (DTW)

Dynamic Time Warping (DTW) is a robust distance measure that incorporates time warping to align two sequences of unequal lengths as shown in Figure 5.2 (b). It allows for the elastic shifting of the x-axis in order to detect similar shapes with different phases. The algorithm finds dynamically the distance matrix between two time aligned sequences and hence the optimum matching path.

Suppose we have two time series Q (query) and C (candidate), of length *n* and *m* respectively, where:

$Q = q_1,q_2,\ldots,q_i,\ldots,q_n$

$C = c_1,c_2,\ldots,c_j,\ldots,c_m$

$$D(i,j)=d(i,j)+\min\left\{\begin{matrix} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{matrix}\right\} \tag{5.2}$$

where, $d(i,j)$ is Euclidean Distance between Q_i and C_j, $i = 1,2,\ldots\ldots,n$; $j = 1,2,\ldots\ldots,m$

The drawback of point to point error calculation is that it does not produce intuitive results, as it compares samples that might not correspond well. DTW solves this discrepancy by recovering optimal alignments between sample points in the two time series. Also the performance of DTW in the context of evaluating musical ornaments like glide will be interesting to observe.

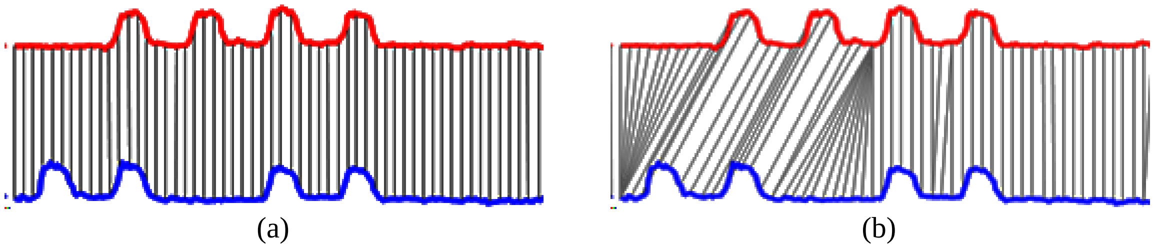

(a) (b)

Figure 5.2 (a) Euclidean Distance of sequences aligned "one to one" (b) "Warped" Time Axis: nonlinear alignments possible

5.2.3. Polynomial Curve Fitting

Whereas the former two distance measures serve to match pitch contours shapes in fine detail, the motivation for the last method is to retain only what may be the perceptually relevant characteristics of the pitch contour. The extent of fit of a 2nd degree polynomial equation to a pitch contour segment has been proposed as a criterion for extracting/detecting *meends* [**9**]. This idea has been extended here to evaluate test singer glides. It was observed in our dataset that 3rd degree polynomial gives a better fit because of the frequent presence of an 'inflection point' in the pitch contours of glides as shown in Figure 5.3. An inflection point is a location on the curve where it switches from a positive radius to negative. The maximum number of inflection points possible in a polynomial curve is *n*-2, where *n* is the degree of the polynomial equation. A 3rd degree polynomial is fitted to the corresponding reference glide, and the normalized approximation error of the test glide with respect to this polynomial is computed. The 3rd degree polynomial curve fit to the reference glide pitch contour will be henceforth referred to as 'model curve'.

An R-Square value measures the closeness of any two datasets. A data set has values y_i each of which has an associated modelled value f_i, then, the total sum of squares is given by,

$$SS_{tot} = \sum_i \left(y_i - \bar{y}\right)^2 \tag{5.3}$$

where, $$\bar{y} = \frac{1}{n}\sum_{i}^{n} y_i \tag{5.4}$$

The sum of squares of residuals is given by,

$$SS_{err} = \sum_{i} \left(y_i - f_i \right)^2 \tag{5.5}$$

and, $$R^2 = 1 - \frac{SS_{err}}{SS_{tot}} \tag{5.6}$$

which is close to 1 if approximation error is close to 0.

In Dataset B, the average of the R-square values of all glides in a song was used to obtain an overall score of the test singer for that particular song.

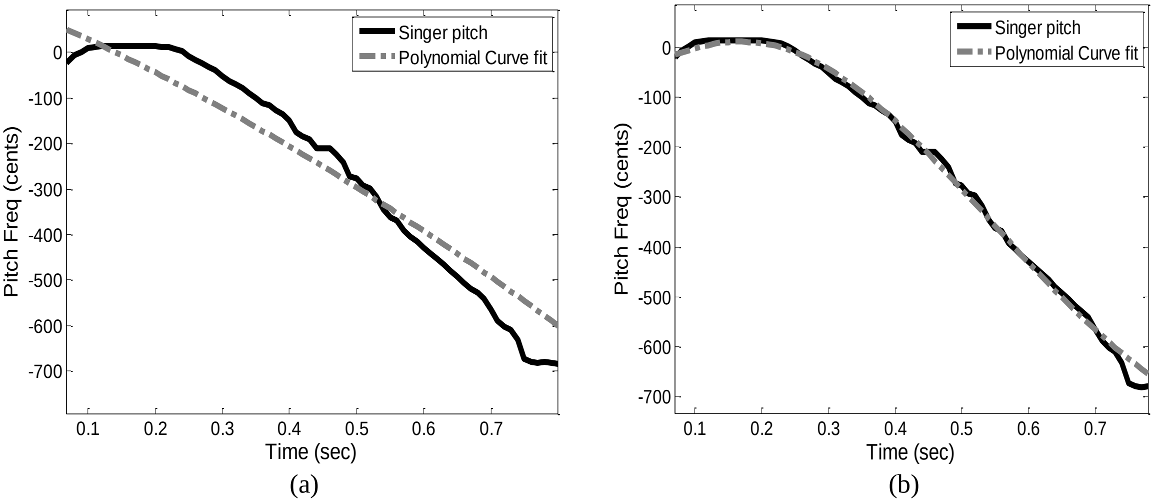


Figure 5.3. Reference glide polynomial fit of (a) degree 2; $P_1(x) = ax^2+bx+c$; R-square = 0.937
(b) degree 3; $P_2(x) = ax^3+bx^2+cx+d$; R-square = 0.989

A slight time shift in rendering the glide by the test singer with respect to the reference singer will result in huge errors even though perceptually this slight time shift might not be penalized. This problem has been taken care of here by calculating the R-square value at time shifts from -100 ms to +300 ms, and selecting the shift that gives the maximum R-square value (minimum error). Larger range of right shift of the test singer's glide has been considered because it was observed that singers tend to render with a time delay more often than being early. As shown in Figure 5.4, the R-square value increases from -0.45 with no shift to 0.904 with 200 ms shift.

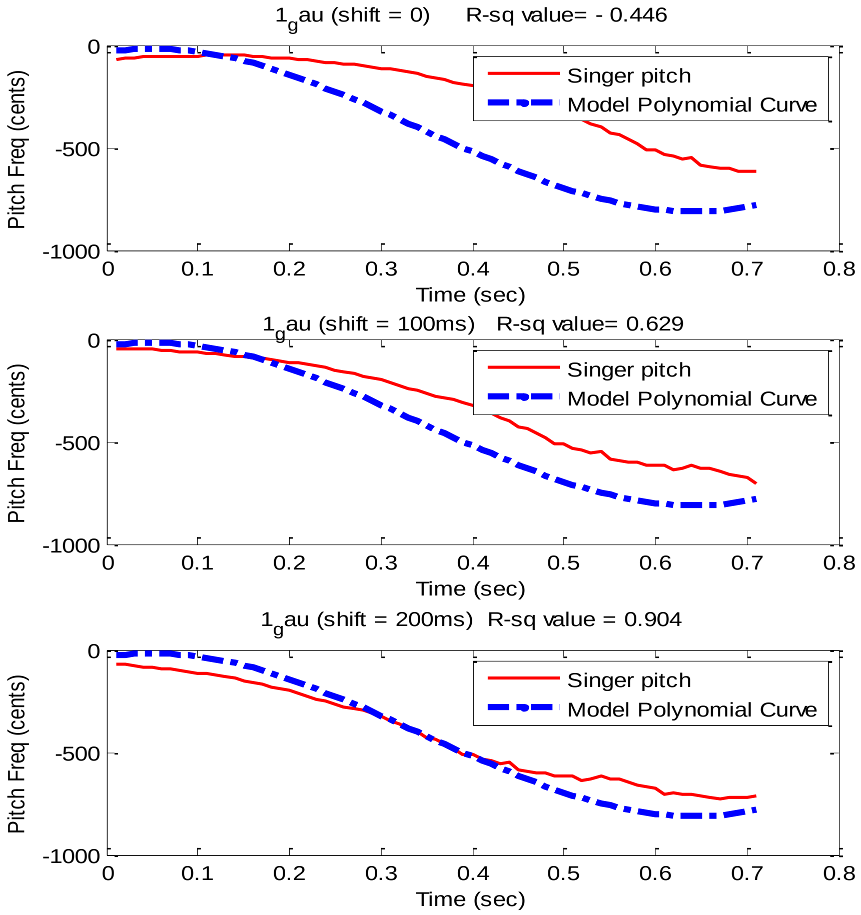


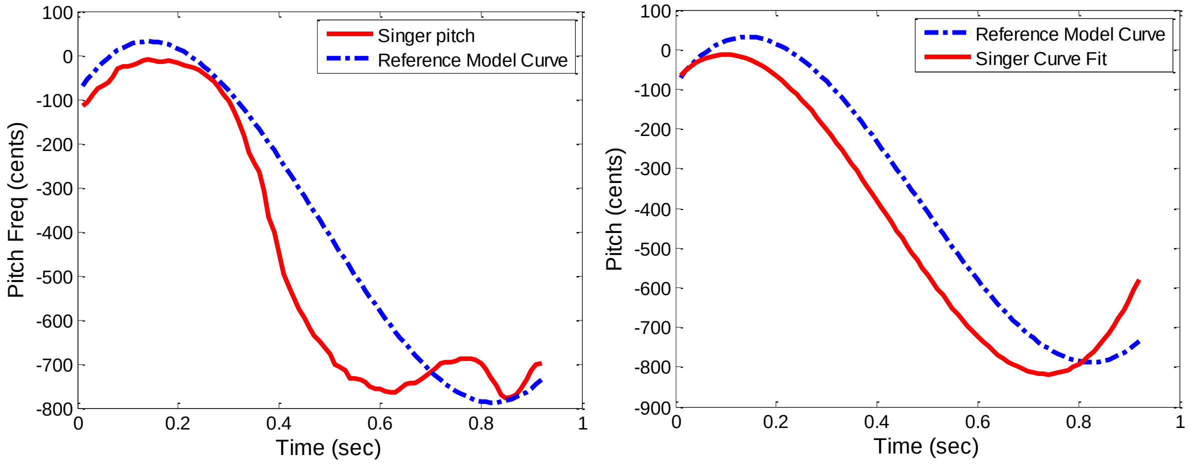
Figure 5.4. Test singer pitch and reference model fit with time shifts of (a) 0 ms, R-sq = -0.446 (b) 100 ms, R-sq = 0.629 and (c) 200 ms, R-sq = 0.904

In this work, three different methods of evaluating a test singer glide based on curve fitting technique have been explored. They are:

i. Approximation error between test singer glide pitch contour and reference model curve (Figure 5.5(a))

ii. Approximation error between test singer glide 3rd degree polynomial curve fit and reference model curve (Figure 5.5(b))

iii. Euclidean distance between the polynomial coefficients of the test glide curve fit and the reference model curve



(a) (b)

Figure 5.5 (a) Test singer pitch contour and reference model curve (b) Test singer polynomial curve fit and reference model curve

5.3. Validation Results and Discussion

A single overall subjective rank is obtained by ordering the test singers as per the *sum* of the individual judge ranks. Spearman Correlation Coefficient (ρ), as explained in Section 4.3. is used to measure correlation between subjective and objective ranks and is employed to validate the system. The results (for Dataset A) appear in Table 5.2.

Table 5.2 Inter-Judges' rank agreement (W) and correlation (ρ) between judges' avg. rank and objective measure rank for the ornament instances for Dataset A. Objective Measure 1: ED, Measure 2: DTW, Measure 3: 3rd degree Polynomial fit with best shift for glide: (i) Test glide pitch contour and model curve (ii) Test glide 3rd deg. polynomial curve fit and model curve (iii) ED between polynomial coefficients of the test glide curve fit and the model curve

Type of Ornament	**Instance no.**	**Inter-judges' rank agreement (W)**	**Correlation between Judges' avg. rank &**				
			Obj. measure 1 rank (ρ)	**Obj. measure 2 rank (ρ)**	**Obj. measure 3 rank (ρ)**		
					(i)	**(ii)**	**(iii)**
Simple Descending Glide	1	0.99	0.85	0.75	0.65	0.65	0.25
	2	0.98	0.95	-0.65	0.15	0.05	0.45
	3	0.82	0.14	0.65	0.66	0.66	0.49
	4	0.87	-0.08	0.08	0.5	0.5	-0.08
	5	0.88	-0.65	0.88	0.94	0.94	1
	6	0.84	0.42	0.57	0.61	0.54	0.07
	7	0.65	0.16	0.16	0.63	0.59	0.76
Complex Descending Glide	1	1	0.2	0.8	0.6	0.6	-0.1
	2	0.95	-0.2	-0.5	0	0.2	0.7
	3	0.96	0.95	0.25	0.65	0.55	0.65
	4	0.73	-0.7	0.63	0.52	0.87	0.4
	5	0.70	0.66	0.87	0.94	0.94	0.87

5.3.1. Dataset A

We observe that out of 12 instances with good inter-judges' agreement (W>0.5), 3rd degree Polynomial Curve fit measure gives the maximum number of instances with a high rank correlation with the judges' rank (ρ >= 0.5). There's a clear trend of Measure 1 (ED) giving poor results as it penalizes for perceptually unimportant factors because a singer may not have sung 'exactly' in the same way as the reference and yet could be correct/ pleasing and scored high by the listeners (Table 5.3). In the case of simple glides, Measure 3 (Polynomial Curve Fit) with methods i. and ii., outperforms Measure 2 (DTW). Both these methods show similar performance, but method i. is computationally less complex. It is observed that all the simple glides that have a good judges' agreement have been characterized very well by Measure 3.

For complex glides, there will be poor curve fit by a low degree polynomial. A lower degree polynomial is able to capture only the overall trend of the complex glide, while the undulations and pauses on intermediate notes that carry significant information about the singing quality (as observed from the subjective ratings) are not appropriately modelled as can be seen in Figure 5.6. So there are chances that a poor singer who replaces the complex glide with a simple glide gets a high rank because the model curve itself does not model reference curve correctly.

Table 5.3 Performance of the different measures for the ornament glide in Dataset A

Measures		**No. of instances that have ρ >= 0.5**	
		Simple Glides (out of 7 with judges' rank agreement)	**Complex Glides (out of 5 with judges' rank agreement)**
1 - Euclidean Distance		2	2
2 - DTW		4	3
3 - 3rd degree Polynomial curve fit	(i)	6	4
	(ii)	6	4
	(iii)	1	2

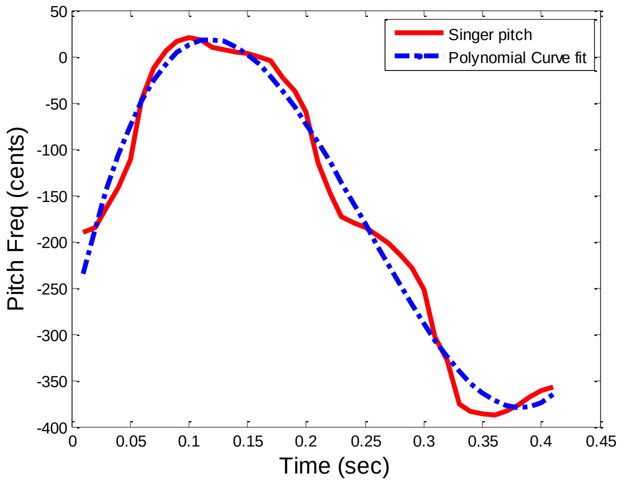


Figure 5.6 Complex glide (reference) modelled by a 3rd degree polynomial

5.3.2. Dataset B

The overall ornament quality evaluation of the singer as evaluated on Dataset B has good inter-judge agreement for almost all singers for both the songs in this dataset. The most frequent rating given by the judges (three out of the four judges) for a singer was taken as the subjective ground truth category for that singer. The cases of contention between the judges (two of the four judges for one class and the other two for another class) have not been considered for objective analysis.

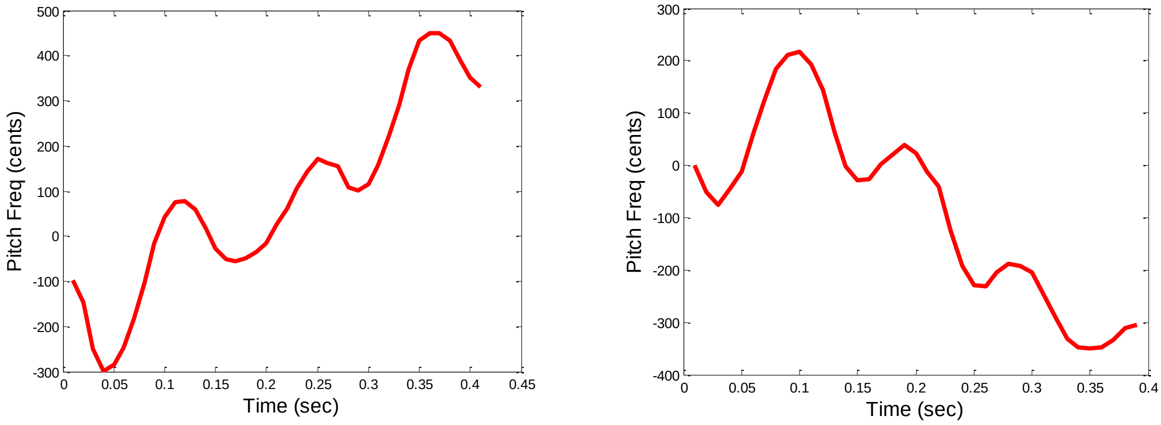
The R-square value of the curve fit measure i. (error between reference model curve and test glide pitch contour) is used for evaluating each of the glide instances for the songs in Dataset B. A threshold of 0.9 was fixed on this measure to state the detection of a particular glide instance. For a test singer, if all the glide instances are detected, the singer's overall objective rating is "good"; if the number of detections is between 75 – 100% of the total number of glide instances in the song, the singer's overall objective rating is "medium"; and if the number of detections is less than 75%, the singer's overall objective rating is "bad". The above settings are empirical. Table 5.4 shows the singer classification confusion matrix. Though no drastic misclassifications between good and bad singer classification is seen but the overall correct classification is very poor 31.25% due to large confusion with the "medium" class. One major reason for this inconsistency was that the full audio clips also contained complex glides and other ornaments that influenced the overall subjective ratings while the objective analysis was based solely on the selected instances of simple glides. This motivates the need of objective analysis of complex ornaments so as to come up with an overall expression rating of a singer.

Table 5.4 Singer classification confusion matrix for Dataset B

Objectively→ **Subjectively↓**	**G**	**M**	**B**
G	0	3	0
M	2	0	4
B	0	2	5

Chapter 6. Assessment of Oscillations-on-Glide

The ornament 'oscillations-on-glide' is typically periodic oscillations riding on a glide-like transition from one note to another. The oscillations may or may not be of uniform amplitude. This ornament has a close resemblance with the ornament *Gamak* in Indian classical music. Some examples of this ornament is shown in Figure 6.1



(a) (b)

Figure 6.1 Fragments of pitch contour extracted from a reference song: (a) ascending glide with oscillations (b) descending glide with oscillations

6.1. Database

This section consists of the reference data, test singing data and the subjective rating description.

6.1.1. Reference and Test Dataset

The reference dataset, consisting of polyphonic audio clips from popular Hindi film songs rich in ornaments, were obtained as presented in Table 6.1. The pitch tracks of the ornament clips were isolated from the songs for use in the objective analysis. Short phrases containing the ornament clips (1 - 4 sec) were used for subjective assessment as described later in this section. The reference songs were sung and recorded by 6 to 11 test singers (Table 6.1).

Table 6.1 ‘Oscillations-on-glide’ database description

Son g No.	Song Name	Singer	No. of ornament clips	No. of Test singers	Total no. of test tokens	Characteristics of the ornaments
1.	Ao Huzoor (Kismat)	Asha Bhonsle	3	6	18	All three instances are descending oscillations-on-glide. Duration: 400 ms (approx.)
2.	Nadiya Kinare (Abhimaan)	Lata Mangeshkar	3	8	24	All three instances are ascending oscillations-on-glide. Duration: 380 - 450 ms (approx.)
3.	Naino Mein Badra (Mera Saaya)	Lata Mangeshkar	13	11	143	All thirteen instances are ascending oscillations-on-glide. Duration: 300 - 500 ms (approx.)

6.1.2. Observations on pitch contour of oscillations-on-glide

This ornament can be typically characterized by the rate of transition, rate of oscillation and amplitude modulation (A.M.). Rate of oscillations is defined as the number of cycles per second. The range of the oscillation rate is seen to be varying from 5 to 11 Hz approximately as observed from the 19 instances of the reference ornament. Some observations for these 19 reference instances are tabulated in Table 6.2. 11 out of the 19 instances are within the vibrato range of frequency, but 8 are beyond the range. Also 7 of the instances show amplitude modulation. The rate of transition varied from 890 to 2000 cents per second.

Table 6.2 Observations on the pitch contour of oscillations-on-glide

Rate range (Hz)	# of instances without A.M.	# of instances with A.M.
5 – 8	5	6
8 – 10	6	0
10 – 12	1	1

6.1.3. Subjective Assessment

6.1.3.1. Holistic Ground Truth

Three human experts were asked to give a holistic categorical rating (Good (G), Medium (M) and Bad (B)) to each ornament instance of the test singers. The most frequent rating given by the judges (two out of the three judges) for an instance was taken as the subjective ground truth category for that ornament instance. Out of the total of 185 test singers’ ornament tokens (as can be seen from Table 6.1), 105 tokens were subjectively annotated and henceforth used in the validation experiments. An equal number of tokens were present in each of these classes (35 each). Henceforth whenever an ornament instance of a singer is referred to as good/medium/bad, it implies the subjective rating of that ornament instance.

6.1.3.2. Parameter-wise Ground Truth

Based on the kind of feedback that may be obtained from a music teacher about the quality of these ornaments, the test singers' a subset of the test ornament tokens (75 test tokens out of 105) were subjectively assessed by one of the judges separately for each of the three parameters – accuracy of the glide (start and end notes, and trend), amplitude of oscillation, and rate (number of oscillations) of oscillation. For each of these parameters, the test singers were categorized into good/medium/bad for each ornament instance. These ratings have been used to assess the performance of each of the corresponding objective attributes.

6.2. Modelling Parameters

From observations, it was found that modelling of this ornament can be divided into 2 components with 3 parameters in all:

i. **Glide**

ii. **Oscillation**

   a. **Amplitude**

   b. **Rate**

Glide represents the overall monotonous trend of the ornament while transiting between two correct notes. **Oscillation** is the pure vibration around the monotonous glide. Large amplitude and high rate of oscillations are typically considered to be good and requiring skill. On the other hand, low amplitudes of oscillation makes the rate of oscillation immaterial, indicating that rate should be evaluated only after the amplitude of oscillation crosses a certain threshold of significance.

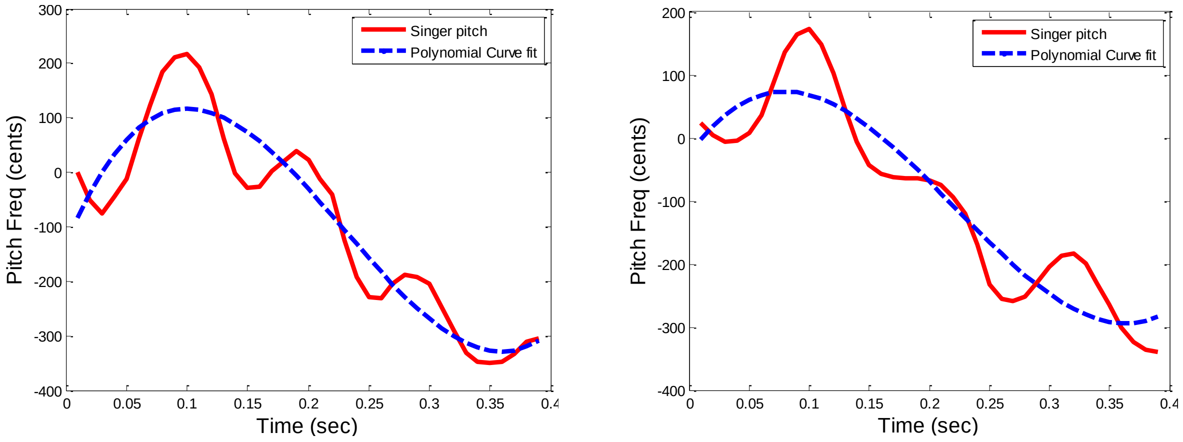
6.3. Implementation of Objective Measures

6.3.1. Glide

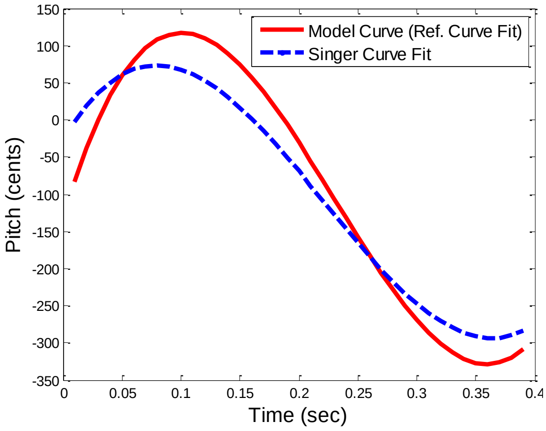
Glide modelling, as explained in Chapter 5. , involves a 3rd degree polynomial approximation of the reference ornament pitch contour that acts as a model curve to evaluate the test ornament. A similar approach has been taken to evaluate the glide parameter of the ornament oscillations-on-glide. The 3rd degree polynomial curve fit is used to capture the trend of the ornament. Out of the two options of polynomial modelling (as explored in Section 5.2.3. of comparing the reference model curve fit with either the test ornament pitch contour or the test ornament 3rd degree curve fit, the latter has been chosen here. This is because a test singer may have replaced the ornament oscillations-on-glide with a simple glide. Comparison of the reference model curve fit with the test pitch contour in that case would give a high score. While a test singer who does the oscillations also, will be penalized as the pitch contour will not exactly fit the reference model curve fit. Since the glide parameter of this ornament characterizes only the trend, and the 3rd degree curve fit captures this trend well, thus the curve fits of the reference and test ornaments are compared to measure the accuracy of this trend. The procedure for evaluating the same is described as follows:

- Fit a “trend model” (3rd degree polynomial curve fit) in the reference ornament (Figure 6.2 (a))
- Similarly fit a 3rd degree curve into the test singer ornament (Figure 6.2(b))
- A measure of distance of the test singer curve fit from the trend model evaluates the overall trend of the test singer’s ornament

R-square value is the distance measure used here; R-sq close to 1 implies closer to the trend model (reference model) (Figure 6.2(c)). This measure is henceforth referred to as *glide feature*.



(a) (b)



(c)

Figure 6.2 (a) 'Trend Model'; 3rd degree curve fit into reference ornament pitch (b) 3rd degree curve fit into test singer ornament pitch (c) Trend Model and Test curve fit shown together; R-square = 0.92

6.3.2. Oscillations

- To analyze only the oscillations of the ornament, we need to first subtract the trend from it. This is done by subtracting the vertical distance of the lowest point of the curve from every point on the pitch contour, and removing DC offset, as shown in Figure 6.3.
- The trend-subtracted oscillations, though appear to be similar to vibrato, they are different in following ways:

  i. Vibrato has approximately constant amplitude across time, while this ornament may have varying amplitude, much like amplitude modulation, and thus frequency domain representation may show double peaks or side humps

  ii. The rate of vibrato is typically between 5 - 8 Hz [**3**], while the rate of this ornament may be as high as 10 Hz

- These oscillations are, by and large, characterized by their amplitude and rate, both of which are explored in frequency as well as time domain.

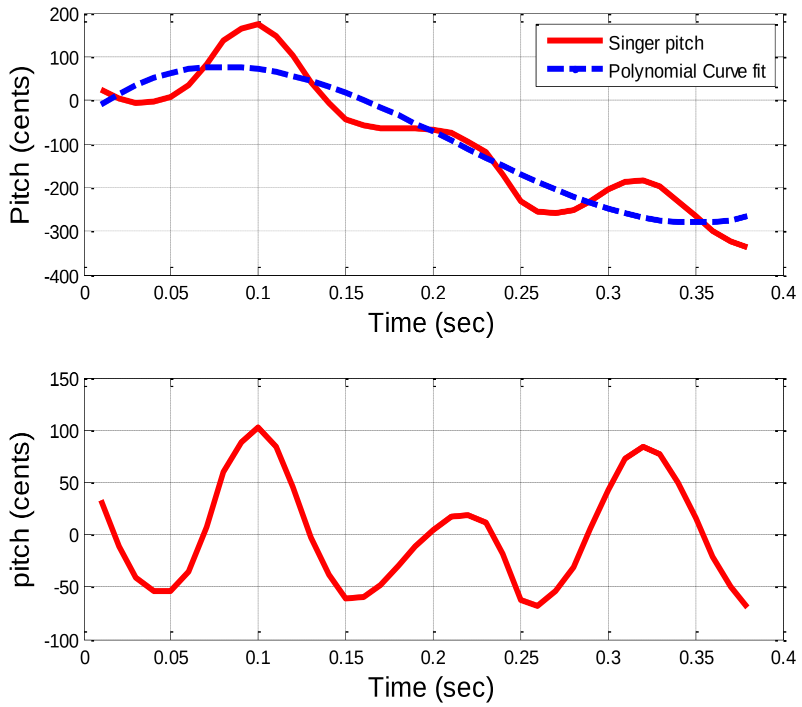


Figure 6.3 Trend Subtraction

6.3.2.1. Frequency domain attributes:

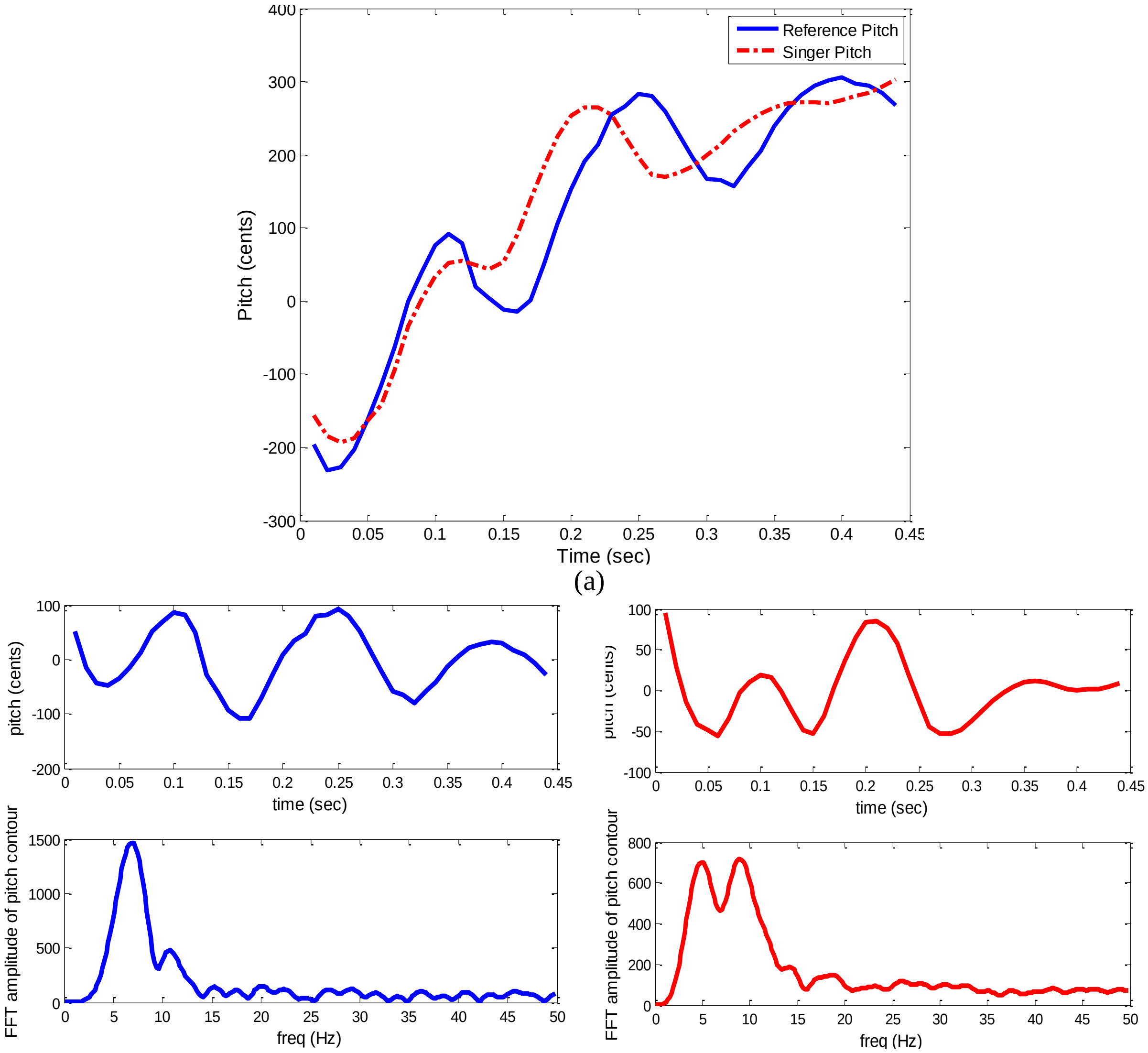
- ***Amplitude*** – Ratio of the peak amplitude in the magnitude spectrum of test singer ornament pitch contour to that of the reference. This measure is henceforth referred to as *frequency domain oscillation amplitude feature (FDOscAmp).*

$$FDOscAmp = \frac{\max\left(\left|Z_{test}\left(k\right)\right|\right)}{\max\left(\left|Z_{ref}\left(k\right)\right|\right)} \tag{6.1}$$

  where $Z_{test}(k)$ and $Z_{ref}(k)$ are the DFT of the mean-subtracted pitch trajectory $z(n)$ of the test singer and reference ornaments respectively.

- ***Rate*** – Ratio of the frequency of the peak in the magnitude spectrum of the test singer ornament pitch contour to that of the reference. This measure is henceforth referred to as *frequency domain oscillation rate feature (FDOscRate).*

The ratio of energy around test peak frequency to energy in 1 to 20 Hz may show spurious results if the test peak gets split into two due to amplitude modulation (Figure 6.4). Also it was observed that amplitude modulation does not affect the assessment of singers perceptually. Thus the scoring system should be designed to be insensitive to amplitude modulation. This is taken care of in frequency domain analysis by computing the sum of the significant peak amplitudes (3 point local maxima with a threshold of 0.5 of the maximum on the magnitude) and average of the corresponding peak frequencies and computing the ratio of these features of the test ornament to that of the reference ornament.



(a)

(b)

(c)

Figure 6.4 (a) Reference and Test ornament pitch contours for a “good” test instance , (b) Trend subtracted reference ornament pitch contour and frequency spectrum, (c) Trend subtracted test singer ornament pitch contour and frequency spectrum

6.3.2.2. Time domain attributes

Due to the sensitivity of frequency domain measurements to the amplitude modulation that may be present in the trend-subtracted oscillations, the possibility of time-domain characterisation is explored. The pitch contour in time domain may sometimes have jaggedness that might affect a time domain feature that uses absolute values of the contour.

Hence a 3-point moving average filter has been used to smoothen the pitch contour (Figure 6.5)

- ***Amplitude*** – Assuming that there exists only one maxima or minima between any two zero crossings of the trend subtracted smoothened pitch contour of the ornament, the amplitude feature computed is the ratio of the average of the highest two amplitudes of the reference ornament to that of the test singer ornament. The average of only the highest two amplitudes as opposed to averaging all the amplitudes has been used here to make the system robust to amplitude modulation (Figure 6.5 Trend subtracted pitch contour and smoothened pitch contour with zero crossings and maxima and minima marked). This measure is henceforth referred to as *time domain oscillation amplitude feature (TDOscAmp).*

- ***Rate*** – The rate feature in time domain is simply the ratio of the number of zero crossings of ornament pitch contour of the test singer to that of the reference (Figure 6.5). This measure is henceforth referred to as *time domain oscillation rate feature (TDOscRate).*

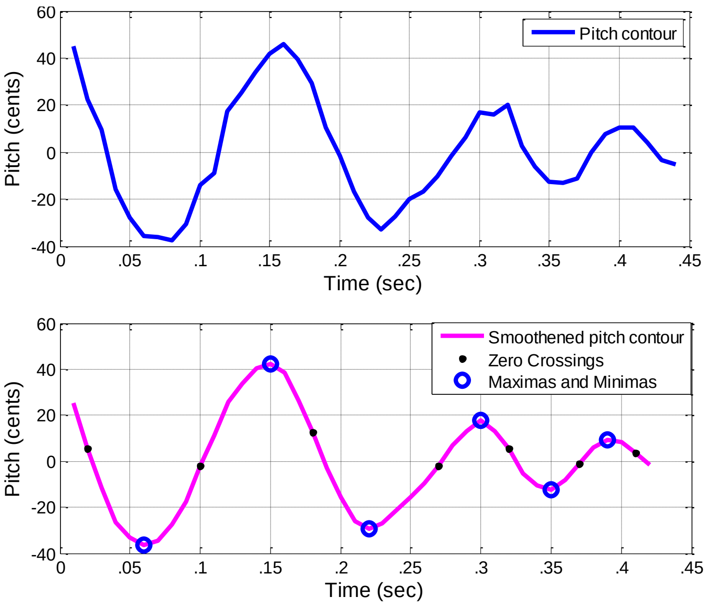


Figure 6.5 Trend subtracted pitch contour and smoothened pitch contour with zero crossings and maxima and minima marked

6.4. Results and Discussion

This section first describes the performance of the different measures of each of the modelling parameters using the **parameter-wise ground truths** for validation. Then the different methods of combining the best attributes of the individual model parameters to get a holistic objective rating of the ornament instance have been discussed.

6.4.1. Glide Measure

In the scatter plot (Figure 6.6), the objective score is the glide measure for each instance of ornament singing that are shape coded by the respective subjective rating of glide (parameter-wise ground-truth). We observe that the “bad” ratings are consistently linked to low values of the objective measure. The “medium” rated tokens show a wide scatter in the objective measure. The medium and the good ratings were perceptually overlapping in a lot of cases (across judges) and thus the overlap shows up in the scatter plot as well. A threshold of **0.4** on the objective measure would clearly demarcate the bad singing from the medium and good singing. It has been observed that even when the oscillations are rendered very nicely, there is a possibility that the glide is bad (Figure 6.7). It will be interesting to see the weights that each of these parameters get in the holistic rating.

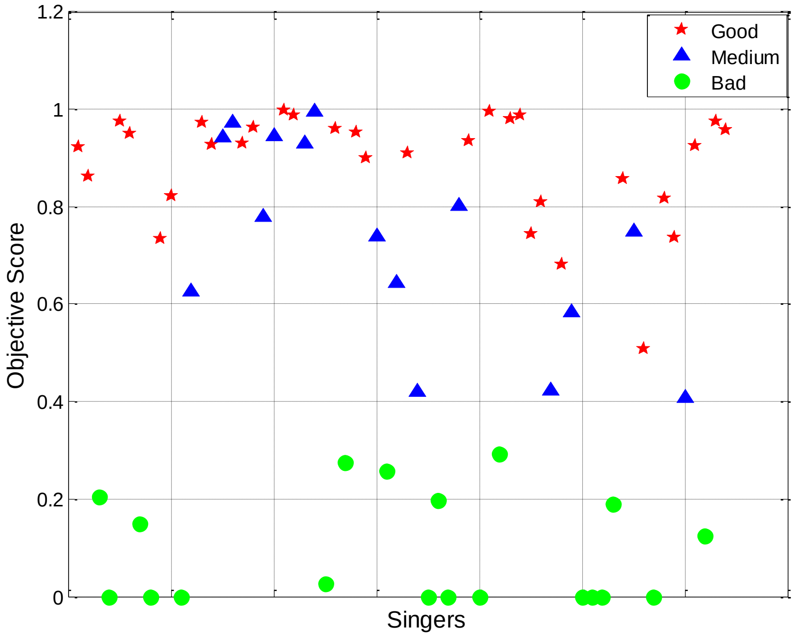


Figure 6.6 Scatter Plot for Glide Measure

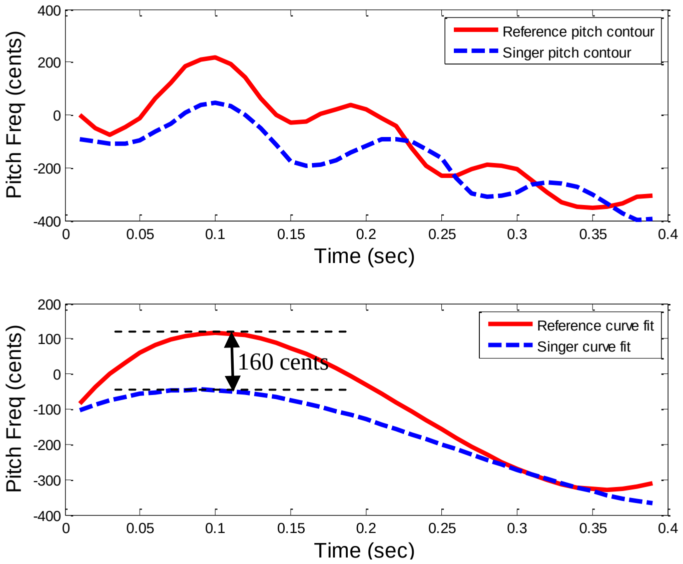
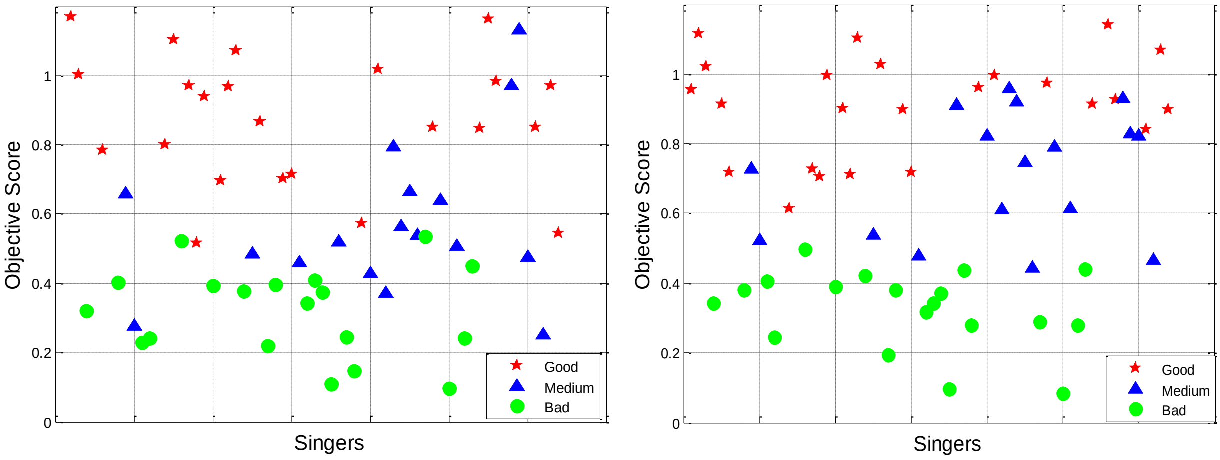


Figure 6.7 Reference and singer ornament pitch contour and glide curve fits

6.4.2. Oscillation Amplitude Measures

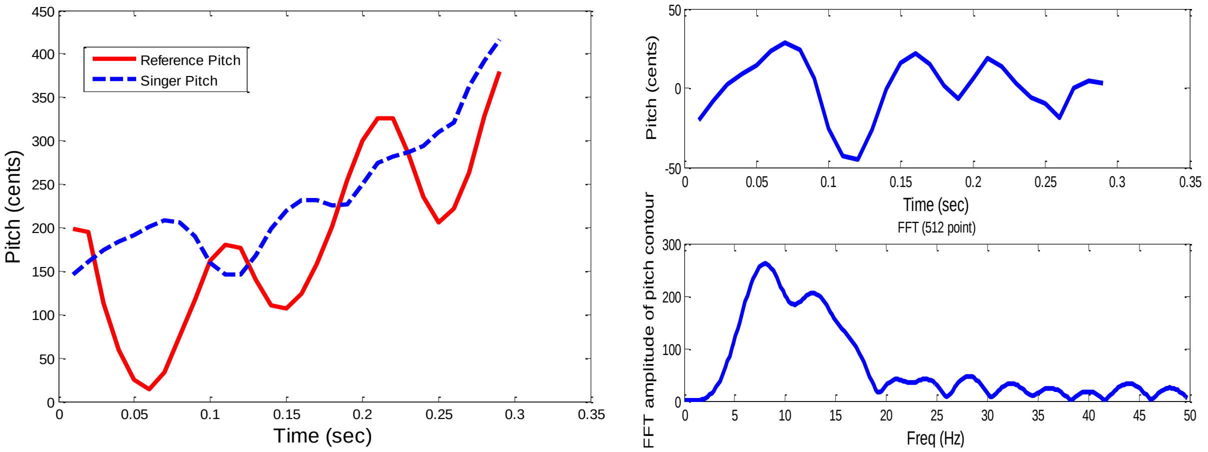
In the scatter plot (Figure 6.8), the objective score is the oscillation amplitude measure for each instance of ornament singing that are shape coded by the respective subjective rating of oscillation amplitude (parameter-wise ground-truth). As seen in the scatter plot, both frequency and time domain features by and large separate the good and the bad instances well. But there are a number of medium to bad misclassification by the frequency domain feature assuming a threshold at objective score equal to 0.4. A number of bad instances are close to the threshold, this happens because of occurrence of multiple local maxima in the spectrum of the bad ornament that add up to have a magnitude comparable to that of the reference magnitude, and hence a high magnitude ratio (Figure 6.9). Also a few of the good instances are very close to this threshold in frequency domain analysis. This happens because of the occurrence of amplitude modulation that reduces the magnitude of the peak in the magnitude spectrum (Figure 6.10).

The number of misclassifications by the time domain amplitude feature is significantly less. The mediums and the goods are clearly demarcated from the bads with a threshold of **0.5** only with a few borderline cases of mediums.



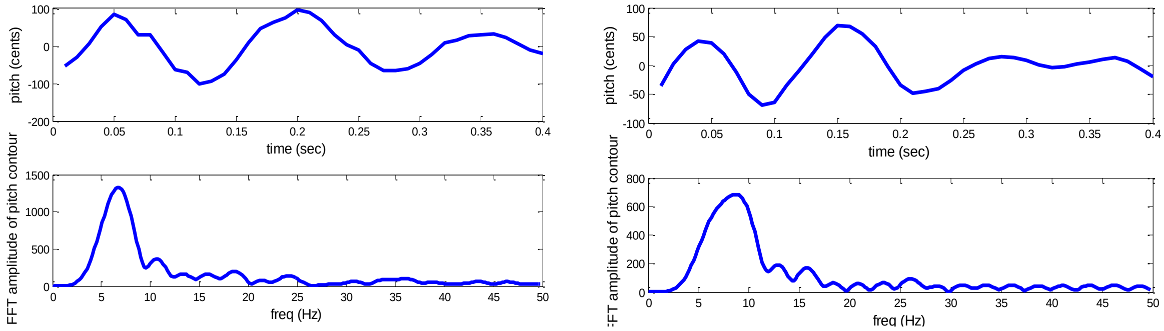
(a) (b)

Figure 6.8 Scatter plot for Oscillation Amplitude measure in (a) Frequency domain (b) Time domain



(a) (b)

Figure 6.9 (a) Bad ornament pitch along with reference ornament pitch (b) Trend subtracted bad ornament pitch from (a) and its magnitude spectrum



(a) (b)

Figure 6.10 Trend subtracted ornament pitch and magnitude spectrum of (a) Reference (b) Good ornament instance

6.4.3. Oscillation Rate measures

It is expected that perceptually low amplitude of oscillation makes the rate of oscillation irrelevant; hence the instances with bad amplitude (that do not cross the threshold) should not be evaluated for rate of oscillation.

It is observed that while there is no clear distinction possible between the three classes when rate of oscillation is analyzed in frequency domain (Figure 6.11 (a)), but interestingly in time domain, all the instances rated as bad for rate of oscillation already get eliminated by the threshold on the amplitude feature and only the mediums and the goods remain for rate evaluation. The time domain rate feature is able to separate the two remaining classes reasonably well with a threshold of **0.75** on the objective score that result in only a few misclassifications.

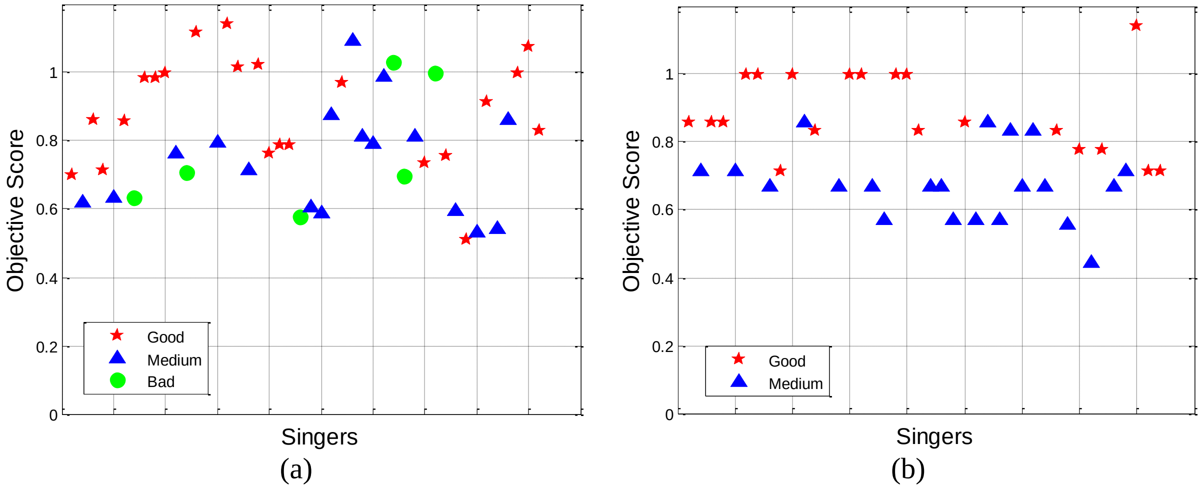


Figure 6.11 Scatter plot for Oscillation Rate measure in (a) Frequency domain (b) Time domain

6.4.4. Obtaining holistic objective ratings

The glide measure gives a good separation between the bads and the goods/mediums and the time domain measures for oscillation amplitude and rate clearly outperform the corresponding frequency domain measures. Thus the **glide measure**, **TDOscAmp** and **TDOscRate** are the three attributes that will be henceforth used in the experiments to obtain holistic objective ratings.

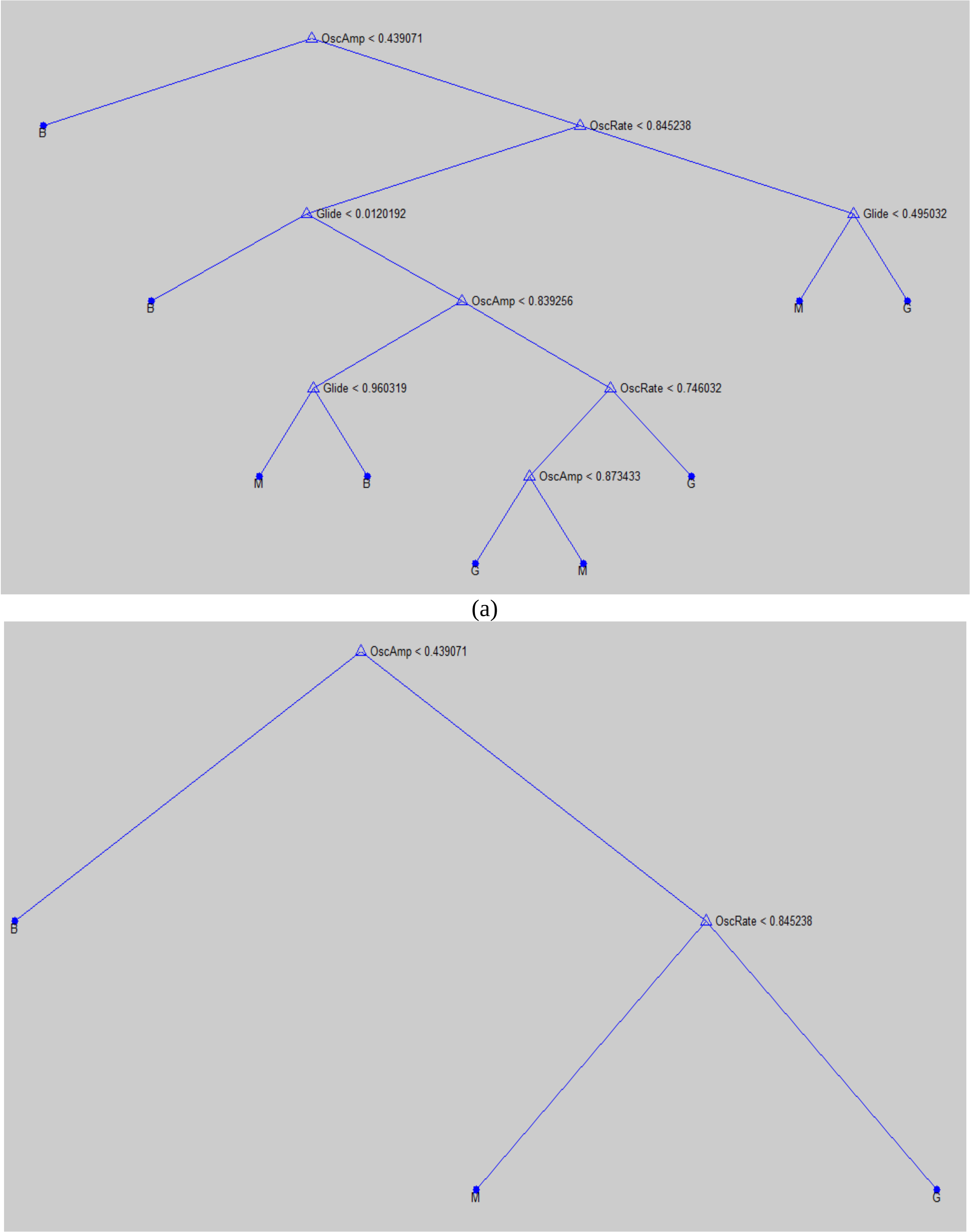
A 7-fold cross-validation classification experiment is carried out for the 105 test tokens with the holistic ground truths. In each fold, there are 90 tokens in train and 15 in test. Equal distribution of tokens exists across all the three classes in both train and test sets. Two methods of obtaining the holistic scores have been explored – complete machine learning, and knowledge-based machine learning as explained in the following subsections.

6.4.4.1. Machine Learning

One method for obtaining holistic objective ratings that has been explored in this work is to train a classification decision tree using the machine learning tool, **C**lassification **a**nd **R**egression **T**ree (CART) (as provided by The MATLAB Statistics Toolbox) in 7-fold cross-validation mode to train a classification decision tree. The tool has been used to train the decision tree for predicting the subjective ratings as a function of the three attributes. Testing in each of the folds has been done once with the full tree and next with the pruned tree. The 'treeprune' command does a 10-fold cross-validation on the training data to estimate the best level of pruning. As mentioned in the help menu of MATLAB, the best level of pruning is the one that produces the smallest tree that is within one standard error of the minimum-cost subtree. Full tree cross-validation experiment gives 30.48% misclassification while pruned tree gives 27.6 – 30.48% misclassifications. While estimating the best level of pruning, internal 10-fold cross-validation does different divisions of data every time the program is executed, and hence the number of misclassifications varies for the pruned tree classification. The full tree for the entire dataset (105 tokens) and the corresponding pruned tree are shown in Figure 6.12. One reason for glide feature not appearing in the pruned tree could be the fact that while holistically rating an ornament, the listeners gave more weightage to the quality of its oscillation parameter than that of its glide parameter. Table 6.3 shows the results of cross-validation experiment with full tree.

Table 6.3 Token classification results of 7-fold cross-validation with full tree (CART) using 3 attributes

Objectively→ **Subjectively↓**	**G**	**M**	**B**
G	**23**	12	0
M	8	**21**	6
B	0	6	**29**



(a)

(b)

Figure 6.12 (a) Full classification tree (b) corresponding pruned tree

6.4.4.2. Knowledge-based Thresholds

From the thresholds derived from the observations of the scatter plots (Section 6.4.1. 6.4.2. 6.4.3. and combining the two time domain features for oscillation using the parameter-wise ground-truths, as explained earlier, we finally have two attributes – the glide measure and the combined oscillation measure. Glide measure gives a binary decision (0, 1) while the combined oscillation measure (TDOsc) gives a three level decision (0, 0.5, 1). Using the manual knowledge-based thresholds obtained, we have a decision tree representation for each of these features as shown in Figure 6.13. Each branch in the tree is labeled with its decision rule, and each terminal node is labeled with the predicted value for that node. For each branch node, the left child node corresponds to the points that satisfy the condition, and the right child node corresponds to the points that do not satisfy the condition. With these decision boundaries, the performance of the individual attributes is shown in Table 6.4.

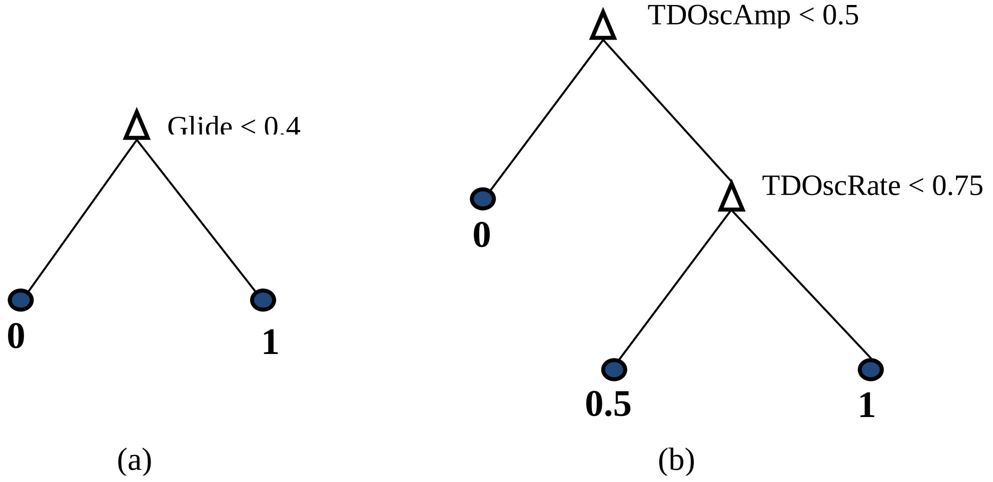


Figure 6.13 Knowledge-based thresholds on the features (a) Glide (b) Oscillation

Table 6.4 Summary of performance of the chosen attributes with knowledge-based thresholds and parameter-wise ground-truths

Attribute →	**Glide Measure**		**TDOsc Measure**		
Threshold→ **Subjective Category↓**	**1**	**0**	**1**	**0.5**	**0**
G	41	0	0	5	28
M	15	0	3	16	4
B	0	19	19	0	0

Once the knowledge-based thresholds are applied to the features to generate the quantized and simplified features **Glide Measure** and **TDOsc Measure**, the task of combining these two features, for a holistic rating of an ornament instance has been carried out by two methods:

i. Linear Combination

In each fold of the 7-fold cross-validation experiment, this method searches for the best weights for linearly combining the two features (glide measure and TDOsc measure) on the train dataset by finding the weights that maximizes the correlation of the objective score with the subjective ratings.

The linear combination of the features is given by

$$h = w_1 g + (1 - w_1) o \tag{6.2}$$

where w_1 and (1 - w_1) are the weights, g and o are the glide and oscillation features respectively and h is the holistic objective score. The holistic subjective ratings are converted into three numeric values (1, 0.5, 0) corresponding to the three categories (G, M, B). The correlation between the holistic objective scores and numeric subjective ratings is given by

$$corr = \frac{\sum_i \left(h_i \cdot GT_i\right)}{\sqrt{\sum_i h_i^2 \sum_i GT_i^2}} \tag{6.3}$$

where h_i and GT_i are the holistic objective score and numeric holistic ground truth (subjective rating) of an ornament token i. Maximizing this correlation over w_1 for the train dataset gives the values of the weights for the two features.

The glide attribute got a low weighting (0.15 – 0.19) as compared to that of the oscillation attribute (0.85 – 0.81). The final objective scores obtained using these weights on the test data features lie between 0 and 1 but are continuous values. However, clear thresholds are observed between good, medium, and bad tokens as given in Table 6.5. With these thresholds, the 7-fold cross-validation experiment gives 22.8% misclassification. The performance of the linear combination method is shown in Table 6.6.

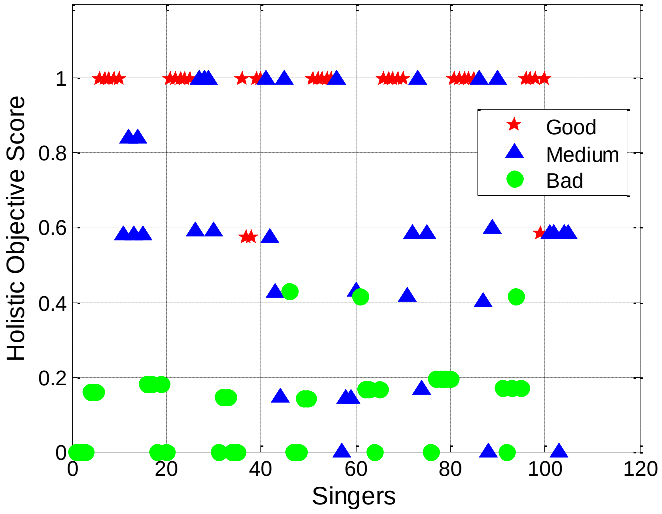


Figure 6.14 Scatter plot of the holistic objective score obtained from Linear Combination method

Table 6.5 Thresholds for objective classification on holistic objective score obtained from Linear Combination method

Holistic Objective Score	**Objective classification**
>= 0.8	G
0.35 – 0.8	M
<0.35	B

Table 6.6 Token classification results of 7-fold cross-validation with Linear Combination method

Objectively→ **Subjectively↓**	**G**	**M**	**B**
G	**32**	3	0
M	11	**17**	7
B	0	3	**32**

ii. Decision boundaries using CART

Another method of obtaining a holistic objective rating of an ornament instance is to obtain decision boundaries from classification tree trained on the two knowledge-based quantized features Glide measure and TDOsc measure. The 7-fold cross-validation experiment has been carried out and testing in each of the folds has been done once with the full tree and next with the pruned tree. Both full and pruned tree cross-validation experiments gave 22.8% misclassifications. A full tree for the entire dataset (105 tokens) is shown in Figure 6.15. Because of the simplified nature of the features, the full tree itself is a short tree with a few nodes and branches and hence mostly the best level of pruning comes out to be zero implying that the tree remains un-pruned and thus no difference in performance. Also it is observed that misclassification rate in this case is same as that in linear combination. The token classification confusion matrix is also same for both the cases (Table 6.6). This suggests that the methods of weight space searching and decision tree training are both similar in nature and perform similarly. This even confirms the final thresholds on the holistic objective score in linear combination method (Table 6.5) as the results obtained are same as this automatic decision boundary generating method.

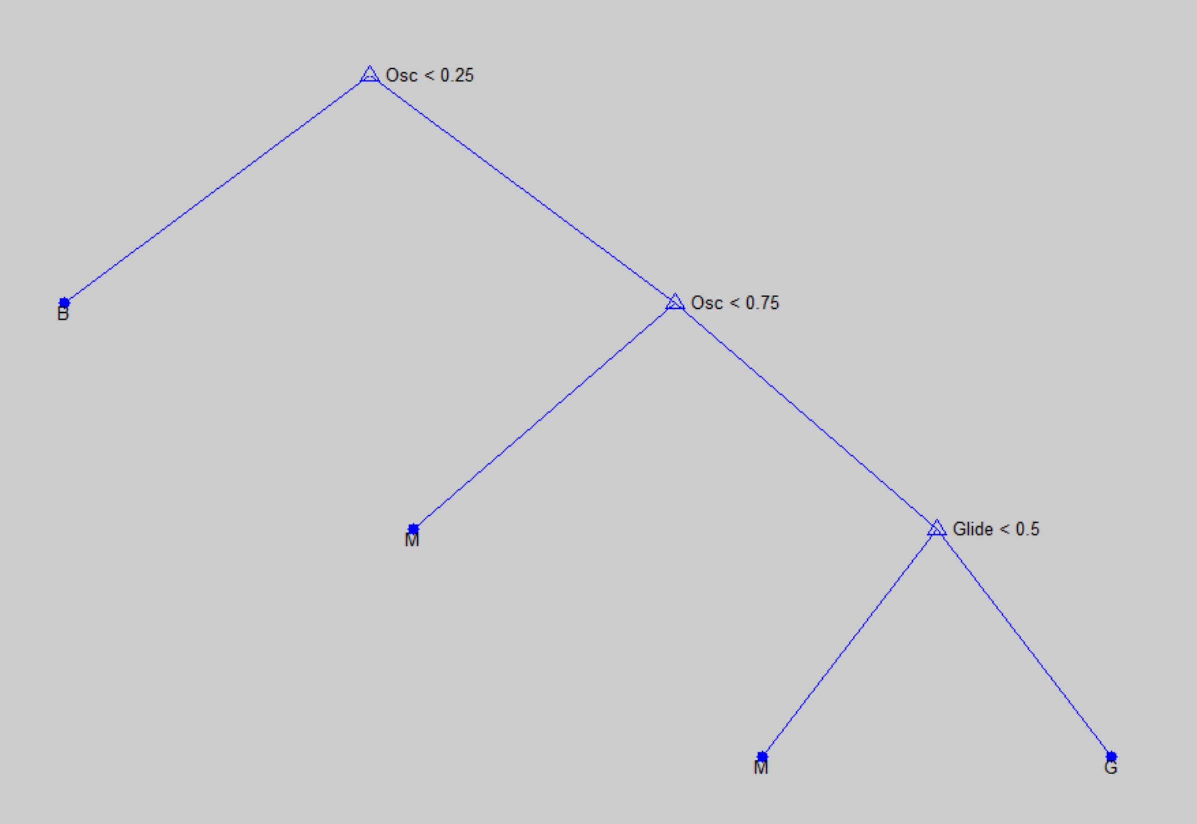


Figure 6.15 Full tree by machine learning using knowledge-based thresholded features

6.4.5. Discussion

- Knowledge-based approach performs better than only machine learning method of classification. The machine learning only method takes real numbers as features and the full tree generated is very complex and given the limited training data, the tree becomes over-trained. The knowledge-based method reduces real numbers into discrete and fewer number of levels and hence when the tree is trained on these simplified features, the full tree generated is fairly simple (lesser nodes). The improvement in performance suggests strongly that the human perceptual judgment has some coarse demarcation factor and fine differences in real numbers are perceptually unimportant.

- The parameters identified for modelling this ornament as well as the features chosen to describe these parameters seem to be appropriate as they correspond fairly well with the subjective judgment.

- The time domain features for oscillation perform better than the corresponding frequency domain features because the system needs to be robust to amplitude modulation as subjective ratings are not affected by it. This robustness is more reliably derived from the time domain features as compared to the frequency domain features.

Chapter 7. Conclusion and Future work

7.1. Conclusion

Modelling and curve fitting methods for assessing ornaments have given encouraging results. Out of 7 simple glides (that closely resemble the Indian classical ornament *Meend*), the objective ratings obtained from 3rd degree polynomial curve approximation method for 6 of them show good correlation with the subjective ratings. The ambitious attempt of modelling and evaluating a complex ornament ‘oscillation on glide’ that closely resembles the Indian classical ornament *Gamak* has been fairly successful (23% misclassifications in 3-category rating). Further, there were no confusions between the two extreme categories. Since this ornament is a major giveaway or differentiator between a good and a bad singer, a fair enough automatic assessment of this ornament will be very useful in singing scoring systems.

Also an attempt was made to get an overall judgment of a singer’s ornamentation skills from the complete audio clip (not just the individual instances) based on objectively evaluated vibratos and glides of the audio clip. This too gave encouraging results clearly indicating the feasibility of objective assessment of singers based on their ornamentation skills.

Various frequency and time domain features have been explored for the vibrato and oscillation modelling parameters. Also polynomial curve approximation methods as well as coefficient space comparisons for modelling the glides and the trends have been explored. For modeling further complex ornaments, good performance is expected by using methods like comparison in higher degree curve fits and curve approximation methods like Bezier splines, B- Splines etc.

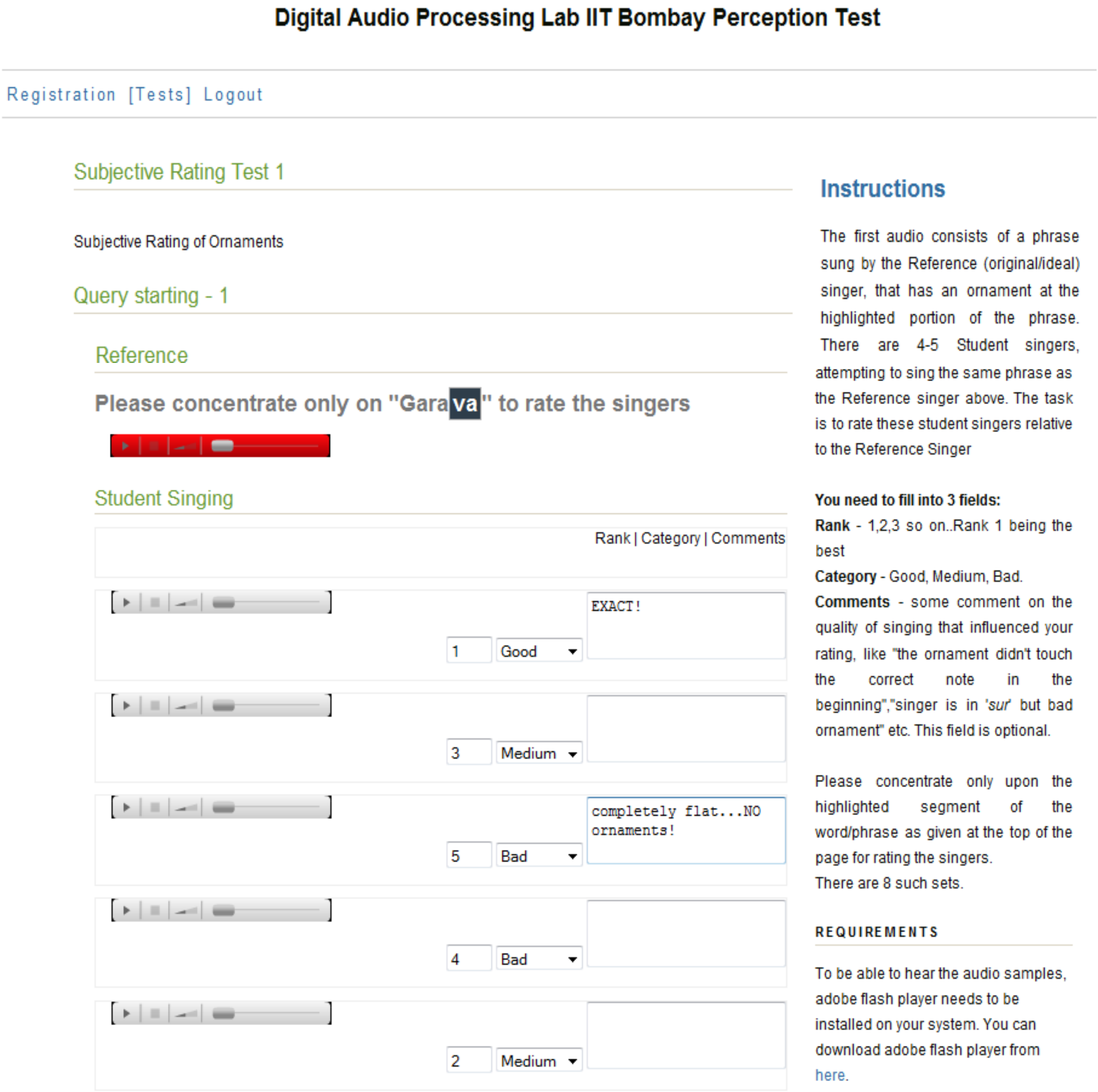
7.2. Future work

In synthesized singing framework in Western music, there has been past work characterizing the behavior of vibrato in the vicinity of portamento pitch transitions (short slide from one note to another) [**14**]. Even though the rate of oscillations during transitions have been observed to be well within typical vibrato rates (which is different from the rate of oscillation on glide observed in Indian singing context) and also the occurrence of oscillations during transition is observed to be less in number, it would be interesting to investigate the relevance of their methods in the Indian context.

Future work will target a framework more suited to Indian classical vocal music where the test singer's rendition may not be time aligned with that of the ideal singer. An ornament assessment system in such a scenario demands reliable automatic detection of ornaments. In the context of purely improvised Indian classical music, the task of evaluation becomes even more challenging as it demands evaluation without a copycat reference and hence the need of more universal ornament models.

Appendix A: Perceptual Test User Interface

An online user graphical interface was designed that had the provision of playback of the ornament audio clips of the reference and the test singers. Human experts were asked to rank (1, 2 and so on) and rate (Good, Medium or Bad) the test singer ornaments clips based on closeness to the reference and specify comments if any. The audio clips were of approximately 3 - 4 secs that comprised of a complete phrase containing the ornament in a word or syllable in that phrase. The ornamented syllable or word to be concentrated upon by the listener is highlighted in the text. The link for the online perceptual test interface is http://www.ee.iitb.ac.in/daplab/subjectivetest/



Digital Audio Processing Lab IIT Bombay Perception Test

Registration [Tests] Logout

Subjective Rating Test 1

Subjective Rating of Ornaments

Query starting - 1

Reference

Please concentrate only on "Garava" to rate the singers

Student Singing

Rank	Category	Comments
1	Good	EXACT!
3	Medium	
5	Bad	completely flat...NO ornaments!
4	Bad	
2	Medium	

Instructions

The first audio consists of a phrase sung by the Reference (original/ideal) singer, that has an ornament at the highlighted portion of the phrase. There are 4-5 Student singers, attempting to sing the same phrase as the Reference singer above. The task is to rate these student singers relative to the Reference Singer

You need to fill into 3 fields:
Rank - 1,2,3 so on..Rank 1 being the best
Category - Good, Medium, Bad.
Comments - some comment on the quality of singing that influenced your rating, like "the ornament didn't touch the correct note in the beginning","singer is in '*sur*' but bad ornament" etc. This field is optional.

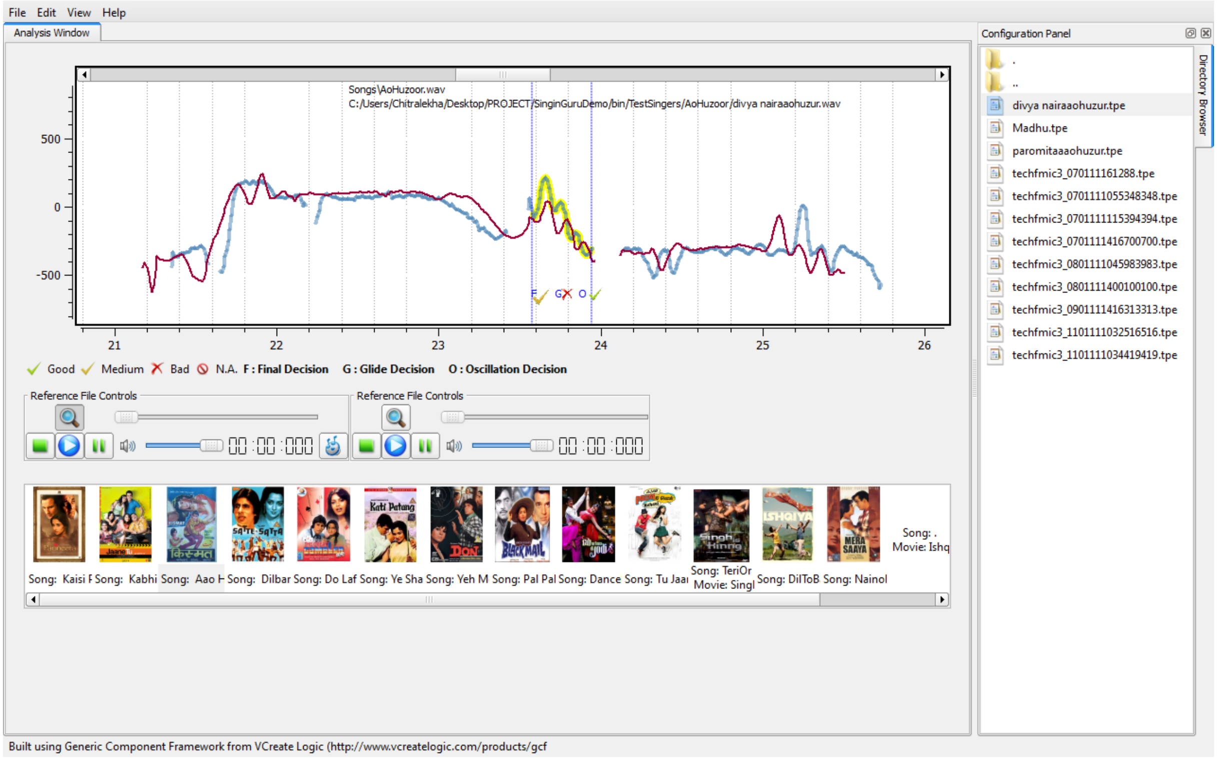
Please concentrate only upon the highlighted segment of the word/phrase as given at the top of the page for rating the singers.
There are 8 such sets.

REQUIREMENTS

To be able to hear the audio samples, adobe flash player needs to be installed on your system. You can download adobe flash player from here.

Appendix B: Oscillations-on-Glide Scoring Interface

For the purpose demonstration, a scoring interface has been built to assess the ornament oscillations-on-glide. It has been provided with the option of loading a reference song and pitch contour and a corresponding test singer pitch contour. The reference pitch is in blue and that of the test singer is in red. The ornaments are marked and pre-stored as metadata and they appear highlighted in yellow on the reference pitch contour. As soon as the test singer pitch contour is loaded, the back-end evaluates the test ornament segments and gives three decision outputs – Glide (G), Oscillation (O) and Final Decision (F), each of which gives a three level decision – Good (a tick mark in green), Medium (a tick mark in yellow) and Bad (a cross in red). If the number of pitches found in the test ornament segment is not equal to that in the reference, then the segment is not evaluated and is marked with a N.A. (shown with a null sign in red). The interface also has the facility of zooming in and out, reference and test singer individual playback and also playback in sync.


File Edit View Help
Analysis Window
500
0
-500
21
22
23
24
25
26
F ✓ G ✗ O ✓
✓ Good ✓ Medium ✗ Bad ⃠ N.A. F : Final Decision G : Glide Decision O : Oscillation Decision
Reference File Controls
00 :00 :000
Reference File Controls
00 :00 :000
Song: Kaisi F Song: Kabhi Song: Aao H Song: Dilbar Song: Do Laf Song: Ye Sha Song: Yeh M Song: Pal Pal Song: Dance Song: Tu Jaa Song: TeriOr Movie: Singl Song: DilToB Song: Nainol Song: . Movie: Ishq
Configuration Panel
Directory Browser

Songs\AoHuzoor.wav
C:/Users/Chitralekha/Desktop/PROJECT/SinginGuruDemo/bin/TestSingers/AoHuzoor/divya nairaaohuzur.wav

.
..
divya nairaaohuzur.tpe
Madhu.tpe
paromitaaaohuzur.tpe
techfmic3_070111161288.tpe
techfmic3_0701111055348348.tpe
techfmic3_0701111115394394.tpe
techfmic3_0701111416700700.tpe
techfmic3_0801111045983983.tpe
techfmic3_0801111400100100.tpe
techfmic3_0901111416313313.tpe
techfmic3_1101111032516516.tpe
techfmic3_1101111034419419.tpe

Built using Generic Component Framework from VCreate Logic (http://www.vcreatelogic.com/products/gcf

References

[1] J. Sundberg, *The science of the singing voice*. Illinois, USA: Northern Illinois Univ. Press, 1987.

[2] N. Amir, O. Michaeli, and O. Amir, "Acoustic and perceptual assessment of vibrato quality of singing students," in *Biomedical Signal Processing and Control*, vol. I, 2006, pp. 144-150.

[3] T. Nakano, M. Goto, and Y. Hiraga, "An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features," in *Interspeech 2006*, Pittsburgh, 2006.

[4] N. Amir, T. Erlich, N. Grabstein, and J. Fainguelernt, "Automated evaluation of singers' vibrato through time and frequenccy analysis of the pitch contour using DSK6713," in *16th International Conference on Digital Signal Processing*, Santorini-Hellas, 2009.

[5] ITC Sangeet Research Academy: A trust promoted by ITC Limited. [Online]. http://www.itcsra.org/alankar/alankar.html

[6] J. Bor, S. Rao, W. Meer, and J. Harvey, *The Raga Guide, A survey of 74 Hindustani Ragas*.: Wyastone Estate Limited, 2002.

[7] A. Datta, R. Sengupta, and N. Dey, "On the possibility of objective assessment of students of Hindustani Music," *Ninaad Journal of ITC Sangeet Research Academy*, vol. 23, pp. 44-57, December 2009.

[8] A. Dutta, R. Sengupta, N. Dey, D. Nag, and A. Mukherjee, "Perceptual evaluation of synthesized ‘meends’ in Hindustani music," in *Frontiers of Research on Speech and Music*, 2007.

[9] A. Datta, R. Sengupta, N. Dey, and D. Nag, "A methodology for automatic extraction of 'meend' from the performances in Hindustani vocal music," *Ninaad Journal of ITC Sangeet Research Academy*, vol. 21, pp. 24-31, December 2007.

[10] A. Datta, R. Sengupta, N. Dey, and D. Nag, "Automatic classification of 'meend' extracted from the performances in Hindustani vocal music," in *Frontiers of Research on Speech and Music*, Kolkata, 2008.

[11] S. Pant, V. Rao, and P. Rao, "A melody detection user interface for polyphonic music," in *NCC 2010*, IIT Madras., 2010.

[12] M.G. Kendall, *Rank Correlation Methods*, 2nd ed. New York: Hafner Publishing Co., 1955.

[13] C. Spearman, "The proof and measurement of association between two things," *Amer. J. Psychol.* , vol. 15, pp. 72-101, 1904.

[14] Maher R. C., "Control of synthesized vibrato during portamento musical pitch transitions," *J. Audio Eng. Soc.*, vol. 56, pp. 18-27, 18-27, 2008.

List of Publications

[1] V. Rao, C. Gupta, and P. Rao, "Context-aware features for singing voice detection in polyphonic music," in *9th International Workshop on Adaptive Multimedia Retrieval*, Barcelona, July 2011. (submitted)

[2] C. Gupta and P. Rao, "An objective evaluation tool for ornamentation in singing", *Proc. of International Symposium on Computer Music Modelling and Retrieval (CMMR) and Frontiers of Research on Speech and Music (FRSM),* Bhubaneswar, India, March 2011.

[3] A. Patil, C. Gupta and P. Rao, "Evaluating vowel pronunciation quality: Formant space matching versus ASR confidence scoring", *Proc. of the National Conference on Communications (NCC)*, Chennai, India, January 2010.