Design and development of an Indian classical vocal training tool

Sharma, Shraddha

2019

Sharma, S. (2019). Design and development of an Indian classical vocal training tool.
Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/136780>

<https://doi.org/10.32657/10356/136780>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0
International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 19:03:21 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**DESIGN AND DEVELOPMENT OF AN INDIAN
CLASSICAL VOCAL TRAINING TOOL**

SHRADDHA SHARMA

**SCHOOL OF ELECTRICAL AND ELECTRONICS
ENGINEERING**

2019

DESIGN AND DEVELOPMENT OF AN INDIAN CLASSICAL VOCAL TRAINING TOOL

SHRADDHA SHARMA

School of Electrical and Electronics Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Master of Engineering

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

28/03/2019

.....
Date

.....
Shraddha Sharma

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

28/03/2019

.....
Date

.....
Prof Wen Changyun

Authorship Attribution Statement

This thesis contains material from one paper accepted at conference in which I am listed as an author.

Chapter 2 and 3 is going to publish (in press) as Shraddha Sharma and Meng Joo Er, “Design and Development of an Indian Classical Vocal Training Tool”, 2nd International Conference on Intelligent Autonomous Systems, 2019.

The contribution of the co-authors is as follows:

- Prof Er Meng Joo provided the initial project direction and helped in editing the manuscript drafts.
- I designed the tool and collected all the necessary data for the calibration of the tool.
- The evaluation parameters were defined by me for the analysis of the audio signal. I produced the results and prepared the manuscript drafts.

28/03/2019

.....
Date

.....
Shraddha Sharma

Acknowledgements

I am using this opportunity to express my gratitude to everyone who supported me throughout the Master of Engineering course. I am thankful for their aspiring guidance, invaluable constructive criticism, and friendly advice during the project work. I am sincerely grateful to them for sharing their truthful and illuminating views on a number of issues related to the project.

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Wen Changyun for guiding me in the thesis writing and helping me in all the time of thesis submission. I would also like to thank Prof. Er Meng Joo for providing me this great opportunity to pursue Masters in the prestigious NTU Singapore. I would like to thank him for the continuous support of my Masters' study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and other career matters. I could not have imagined working on this research project without his motivation and appreciation towards my research topic. He gave me the freedom to choose my research topic according to my interest and continuously guided me to work in the right direction.

Besides my supervisor, I would like to thank the rest of my friends and seniors in NTU for their insightful comments and encouragement, but also for the hard question which incited me to widen my research from various perspectives.

Last but not least, I would like to thank my brother who supported me all throughout my research and encouraged me to work in the right direction. My sincere thanks also go to my parents who made it possible for me to grab this opportunity and never missed a chance to encourage me and support me in any manner possible.

Abstract

When learning any form of classical music vocals without a teacher, even subtle deviations can lead to wrong training, which can be difficult to remedy in the future. In this research work, a real-time tool has been developed to deal with this situation by assisting people in learning Indian classical music. This tool will have a set of pre-defined *Swaras*, *Alankaras*, and *Ragas* (Indian classical music concepts). Users can practice any musical piece from this set and the tool will inform them of the mistakes they make, by smartly matching their voice with the dynamically defined pattern. Users are free to sing in any given scale, which they define in the beginning by singing the root note of their preferred scale.

Using the PredominantPitchMelodia algorithm, the tool identifies the pitch values of the musical piece. From these pitch values, the tool identifies the root-note and set it as a reference. This root-note defines the scale of the user's voice. Further, for identifying the simplest basic musical piece, the k-means clustering method has been implemented along with some index rearrangement. To identify any general musical piece, moving average, gaussian filter and step-detection algorithm have been implemented along with prominent step-filtering method. Voice stability and Pitch accuracy have been proposed to serve as the evaluation criteria. The functioning of the tool is testified on a varied range of audio samples, namely, male and female voices, Ukulele and *Harmonium* audio samples and existing musical instrument tuning applications like GuitarTuna, Ukulele Tuner, and so on have been used as the ground truth for the verification.

For future work, the just intonation tuning method will be implemented as this method is mostly used in Indian classical music. Also, the tool can be formulated for different cultural music (western music, Chinese music, folk music etc.), medical purposes (brain-computer interface, neural disorders like schizophrenia etc.), meditation and for music composition.

List of Figures

3.1	Pitch values of audio signal (of root note) obtained using PredominantPitchMelodia.	22
3.2	k-means clustering to identify the root note.	23
3.3	Pitch values of Alankara audio signal obtained using PredominantPitchMelodia.	24
3.4	k-means clustering to identify the combination of notes present in <i>Alankara</i>	25
3.5	Clusters of n notes to obtain sequence of notes and their comparison with the ideal range of notes for calculating voice stability and pitch accuracy.	26
3.6	Moving average to smoothen the signal.	27
3.7	Moving average to smoothen the signal.	28
3.8	Implementation of Gaussian Filter to identify the steps.	29
3.9	Implementation of Threshold to identify the prominent steps.	30

List of Tables

4.1	Results showing Voice stability and number of notes in ideal range for 5 subjects of basic alankara	37
4.2	Notes in range for 5 subjects of basic alankara	37
4.3	Pitch Accuracy of 5 subjects in Hz for basic alankara	38
4.4	Results showing Voice stability and number of notes in ideal range for 5 subjects of general alankara	39
4.5	Results showing Voice stability and number of notes in ideal range for instrumental recordings of general alankara	39
4.6	Pitch Accuracy of 5 subjects in Hz for general alankara	41
4.7	Notes in range for 5 subjects of general alankara	42
4.8	Pitch Accuracy of two different scales of Harmonium in Hz	43
4.9	Pitch Accuracy of two different scales of Flute in Hz	44

Contents

Acknowledgements	vii
Abstract	ix
List of Figures	xi
List of Tables	xiii
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	2
1.3 Background	3
1.4 Contributions	5
1.5 Organization of the Thesis	6
2 Literature Review	9
2.1 Related Works	9
2.1.1 Music Theory	11
2.1.2 Vocal Melody Extraction	13
2.1.3 Tonic Identification	14
2.1.4 Pattern Recognition	15
2.1.5 Tuning Method	16
2.1.6 Music Performance assessment	17
2.2 Challenges	19
3 Design and Development	21
3.1 Root-note Identification	21
3.2 Basic Alankara Identification	24
3.3 General Alankara Identification	26

3.3.1	Moving Average	28
3.3.2	Step-detection	29
3.4	Summary	30
4	Results and Discussion	33
4.1	Ideal Range Calculation	33
4.2	Evaluation Criteria	34
4.2.1	Notes in Range	34
4.2.2	Voice Stability	35
4.2.3	Pitch Accuracy	35
4.3	Experimental Results	36
4.3.1	Basic Alankara	36
4.3.2	General alankara	38
4.4	Summary	44
5	Conclusions and Future Extensions	47
5.1	Conclusions	47
5.2	Future Work	48
5.3	Applications	49
	Author's Publication(s)	51
	Bibliography	53

Chapter 1

Introduction

Indian classical music has for the past 50 years become an increasingly international phenomenon. A great number of people have had an urge to learn this form of music but lacked resources that could direct them in the right direction. This problem presents an opportunity for a vocal tutor system that can suggest the user improve their vocal skills and provide a detailed analysis of their vocal ability. This vocal tutor can smartly guide the user in practicing classical music on their own and it can be used for music of different cultures in the future.

1.1 Motivation

Learning any form of classical music without a teacher is very difficult. The subtle deviations in music can lead to wrong training and cause harm to the vocal skills of the person. While practicing any piece of music, a person can make mistakes in a particular note of the song but it becomes difficult to find out the exact note where he is not singing properly without proper guidance. The person might have the intuition that somewhere he is singing out of defined scale but how to correct it without the knowledge that where the mistake is happening.

To resolve this issue, there is a requirement of a tool which can assist people to learn, analyze and evaluate the classical music. This tool should be able to automatically generate the ideal sequence of notes which the user should be singing. It should also point out the note where the user is going out of the defined scale and by what factor he is deviating from the defined

note. This tool should provide a detailed analysis of the user's vocal ability and the guidance to improve it.

This tool can be an exceptional help for music artists and guide them in the composition of music. Also, this tool can be extended for medical purposes to cure neural disorders, meditation and so on. The following section explains the background of Indian classical music as it incorporates the most essential element in the construction of this tool.

1.2 Objectives

The objective of designing this tool is to help people practice the different concepts of Indian classical music on their own. This tool can serve the purpose of assisting people learning Indian classical music. In some instances, people are willing to learn music but could not do so because of a lack of adequate facilities. While in other instances, people join music class but feel the need for guidance while practicing music on their own.

This tool should guide the people in learning music like a home tutor and should adjust the features according to the user's voice. It should provide a detailed analysis of the user's voice and direct them in improving their singing skills. This tool should also present a user with the option of learning musical patterns in Indian classical music and also movie songs based on different categories. The songs can be categorised based on the learning level of the user ie. beginner, intermediate, and advanced.

Also, the tool can be used for other musical forms like western music, Chinese music, and folk music. The approach to learning any music form is almost similar, only the music notation is different. So it can be used for any music form.

This tool can also be extended to have additional features of converting a given piece of music or song into its basic formation in Indian classical music. The basic formation will define the scale and rhythm of the song which will expedite the singing lesson for the user. Also, the evaluation parameters like voice stability and pitch accuracy can be used to provide a detailed analysis of the user's vocal ability.

1.3 Background

This section illustrates the important concepts of Indian classical music. These concepts are the building blocks in the designing of the vocal training tool. The Indian classical music form consists of the following components:

- *Swara*: It is a Sanskrit word that means a note in the western music form. Defined as “*Swayam Ranjayathi Ithi Swaram*” which means *swara* makes the mind rejoice by itself. The term *swara* is formed by the first two letters “*Swayam*” and “*Ranjayathi*”. In Indian classical music form, different notes which form an octave are “Sa”, “Re”, “Ga”, “Ma”, “Pa”, “Dha”, and “Ni”. These notes all together are called “*Saptha Swara*”. It is an assortment of frequencies placed periodically in time in order to create a pattern that is soothing when produced over a physical media.

Among *saptha swara* “Sa” and “Pa” are non changing *swaras*. In one musical scale, there are 12 notes. Out of these 12 notes, 7 specific notes form an octave. These 12 notes are subdivided into 7 major notes and 5 minor notes which are known as “*Shudha*” and “*Komal*” *swaras* respectively in the context of Indian classical music form. “Sa”, “Re”, “Ga”, “Ma”, “Pa”, “Dha”, and “Ni” are called *Shudha swaras*. They are also called *Shadja*, *Rishabh*, *Gandhar*, *Madhyam*, *Pancham*, *Dhaivat*, and *Nishad* respectively.

- *Shadja*: This is the root note which is the reference for the location of all other notes in an octave. The root note, “*Sa*”, in Indian classical music form is the base note which defines the scale in which the user is singing. The root note is the foundation of melodies in Indian classical music. All the tones in the musical progression are always in reference and related to the root note. The accompanying instruments in a given musical piece are also tuned using the root note of the lead performer.
- *Alankaras*: *Alankaras* are different possible combinations of notes or *swaras*. The most basic *alankara* is the order of notes in an octave ie. “Sa Re Ga Ma Pa Dha Ni SA”. Likewise, there are several possible combinations of these notes.
- *Raga*: A *raga* consists of at least five notes, and each *raga* provides the musician with a musical framework within which to improvise.

- *Sthayi*: A group of *swaras* from *Sa* to *Ni* are called one *sthayi* (Octave).

The major vocal forms or styles associated with Indian classical music are *Dhrupad*, *Khyal*, and *Tarana*. *Dhrupad* lies in the category of the old style of singing. Traditionally, *Dhrupad* style form is performed by male singers with musical instruments, namely, *tambura* and *pakhawaj*. *Khyal* is the modern vocal form of Indian classical music. *Khyal* contains a greater variety of embellishments and ornamentations compared to *dhrupad*. Another vocal form, *taranas* are medium to fast-paced songs that are used to convey a mood of elation and are usually performed towards the end of the concert.

Although Indian classical music is focussed on vocal performance, instrumental forms have existed since ancient times. In some parts of the world, the instrumental form of Indian music is more popular than vocal form, probably because of the language barrier. One of the instrumental forms of Indian classical music is “*jugalbandi*”. The word *jugalbandi* means, literally, “entwined twins”. This form features a duet of two solo musicians. It is similar to a playful music competition between two musicians playing different musical instruments, namely, *tabla*, *guitar*, *drums*, and so on. *Jugalbandi* is usually performed by two soloists on an equal footing .

In this form, the performer improvises an instrumental musical piece. In *jugalbandi*, both musicians act as lead performers. While any Indian music performance may feature two musicians, a performance can only be deemed a *jugalbandi* if neither is clearly the soloist and nor clearly the accompanist. One of the most popular performers in *jugalbandi* is Ustad Zakir Hussain, the Indian *tabla* virtuoso.

There are various musical instruments associated with Indian classical music form. Some of the popular musical instruments are *Veena*, *Tabla*, *Harmonium*, *Bansuri*, *Shehnai*, *Sarangi*, and so on.

Harmonium is mostly used in teaching vocal skills in Indian classical music. The reason behind this is that *harmonium* can be closely associated with vocal frequency as compared to other musical instruments. *Harmonium* is a free-standing keyboard instrument in which the sound is produced by air, supplied by hand-operated bellows. This instrument has a total of 36 keys and each key denotes a specific *swara*. The combination of these 36 keys forms various scales of Indian classical

music, each scale having 12 keys.

In the vocal sessions of Indian classical music, the teacher determines one scale from several scales to make students practice the opening lessons of Indian classical music. This scale is usually accepted based on the vocal frequency of the student. The vocal frequency is the frequency of the speech of the individual which is generated whenever the person vocalizes. This frequency is unique for each individual.

In the first lesson of the vocal classes of Indian classical music, students are taught to sing the *Shadja swara* ie. *Sa* note. After learning to sing the *Shadja*, the student becomes trained to acquire other notes of the scale. Depending on the frequency of the *Shadja swara*, other notes are determined. There is a mathematical relationship between all the musical notes of a scale. The knowledge of mathematical relation is imparted to the students in terms of music theory. The music theory has distinct terms and keywords for describing the mathematical relationship between different musical notes.

The above-mentioned basics of Indian classical music have been used in the designing of this vocal tutor system. The fundamentals of Indian classical music have been coupled with signal processing and machine learning techniques to produce the best results. The comprehensive research has been carried out for the design of this vocal tutor.

1.4 Contributions

In this research work, the method has been proposed to identify the combination of notes present in a vocal recording by extracting the pitch values of the different notes present applying PredominantPitchMelodia algorithm from Essentia[1] library of python and presenting the combination of these pitches in a sequential manner which operates as a tool to supervise the user learning Indian classical music.

Two different methods have been implemented to identify the sequence of notes. Firstly, the primary sequence of notes is detected by implementing the k-means clustering method along with index rearrangement. In the next approach, the moving average, one-dimensional Gaussian filter, and step-detection algorithm have been implemented to classify the general sequence of notes.

This tool will have a pre-defined list of certain *alankaras*, from which the user can select the one they want to practice. The feature that makes it compatible for every user is that the scale of the musical pattern is selected according to the user's voice. The tool initially asks the user to sing the root note “*Sa*” and define the scale in which they are singing. When the user selects the *alankara* from the pre-defined list and starts to sing a particular piece of music, it recognizes the scale in which the user is singing and examines whether the other notes of *alankara* are lying in that scale or not. It not only provides the comparison of notes with the ideal one but also calculates the voice stability and pitch accuracy of the user's voice, which gives the exact location and number of times the user is going out of scale and by what factor.

1.5 Organization of the Thesis

This dissertation shows how a tool has been designed for improving the singing skills of a beginner and focusses on developing this tool for Indian classical music. The organization of this thesis is as follows. Chapter 2 introduces the previous related work done in the field of music performance assessment. Initially, it describes the existing models based on the assessment of music performances and musical games. Then it compares different approaches to retrieve music information from the audio signal and describe the methods to process it for evaluating the singing performance of the user.

Chapter 3 provides a brief description of the design of the tool. Firstly, it describes the identification of fundamental notes, which define the basis of the Indian music form. Then it focusses on methods implemented for identifying the basic Indian music pattern along with the plots depicting the recognised basic singing pattern of Indian music. Next, it demonstrates the method to identify the general singing pattern for Indian music and describe why the different method has been used for the general Indian music pattern instead of the previous one used for the basic Indian music pattern.

Chapter 4 highlights the tuning method for calculating the ideal range of different musical notes and their relevance in computing the singing performance assessment of the user. It also defines different evaluation parameters which are further used to present the results of the singing exercise. These results are then discussed to give the user an idea about their vocal skills in an understandable form.

Chapter 5 gives a summary of this thesis as well as discusses directions for future work.

Chapter 2

Literature Review

This chapter briefly describes the previous research work carried out in the field of music education. Music education has several applications in various fields, namely, singing, music instruments (like guitar, piano, flute, drums, and so on), learning foreign language, medical purposes (like motor impairment, depression, anxiety, and other health issues). Music information retrieval is the research field in which the information is extracted from music like beats, rhythm, melody, pitch, fundamental note, and so on. This information is further analysed using machine learning techniques like classification and regression to get the desired output.

Audio transcription and digitization of musical data is a very interesting and challenging topic. To extract meaningful information from musical data, methods like deep learning, neural network, and so on are also implemented. *Swara* identification is one of the most important steps in the music information retrieval research. The following sections discuss about the *swara* identification and other previous research carried out in the field of music information retrieval.

2.1 Related Works

This section surveys the previous work related to MIR usage in music education. The music education system can be broadly classified into three categories: play-along CDs and instructional videos, video games, and software for music education. The play-along CDs and instructional videos[2] have played an important role in helping the users to practice music at home

and offering the flexibility to practice at their own pace. However, these types of learning materials are focussed on assisting the user in learning musical instruments and does not provide evaluation parameters, which means that the user has to rely on their own perception and assessment[3] and is quite challenging for the beginners. There are various video games available for music education purposes. The controller of these games operates like a musical instrument. For eg. in the case of guitar related video games, the button of the controller corresponds to the string of the guitar. In some video games, the real musical instrument can be connected through USB instead of the controller. Also, there are video games which provide a set of songs for singing and evaluate the voice by comparing with the original track. However, these video games are limited to pop and rock music genre and the user can only practice music from the given list of songs. This means the user has a limited number of options to practice music.

In melody note transcription, Tony is a software which aimed to make the scientific annotation of melodic content, and especially the estimation of note pitches in singing, more efficient. For achieving this, two main objects of interest are the pitch track, which traces the fundamental frequency contours of pitched sounds in smooth, continuous lines, and the note track, a sequence of discrete note events that roughly correspond to notes in a musical score.

In music education software, Songs2See Game is a game software which performs music assessment for various instruments along with voice. Also, it provides the evaluation parameters for the assessment[4] of the performance. However, the software has mainly addressed the characteristics of western music genre. As the studies show, for defining the evaluation parameters, deep neural networks[5] provide a better result than hand-crafted features for music performance assessment. This study also performs the analysis for western music.

There are several automated singing evaluation methods for karaoke systems. One of the methods is based on assessing the performance and rating it based on diverse music characteristics like pitch, volume, timbre, and rhythm. These systems are perceived to produce ratings close to human rating along with lyrics verifying. However, the automated scoring requires the training data. So, this method does not work quite well without reliable training data. For this purpose, training data should be from professional or trained singers.

From previous literature, it has been found that acoustic features combining spectrum envelope and pitch are used with classifiers trained on sung vow-

els for classification of test vowels segmented from the audio of solo singing. Two classifiers namely, Gaussian Mixture Models and Linear Regression have been observed to perform well on sung vowels. However, the system has not been tested for the typical degradations that arise in telephony and consonant errors has not been focussed on.

As mentioned above, most of the research in the development of music education tools is mostly focussed on western music genre and non-western music forms like Indian classical music, Chinese music, folk music and so on are more complex. To the best of our knowledge, there does not exist any music education tool for Indian classical music form.

The following section surveys the previous work related to the main components employed in the designing of this tool ie. identification of tonic pitch from the vocal recordings, pattern recognition, and tuning method.

2.1.1 Music Theory

Music is a form of art. This art form is an amalgamation of different emotions. One can express love, joy, sadness, anxiety, and nearly all emotions through a different musical piece. Almost everyone relishes listening to music. Some people like rock and pop genre of music while others like soothing and relaxing music. Diverse cultures have their style of music.

In Indian music, multiple varieties are present, namely, Rajasthani music, Punjabi music, Folk music, Filmi, classical music, Indian rock, and Indian pop. The classical music tradition in India includes *Hindustani classical* and *Carnatic classical*. These two foremost music traditions have developed over numerous years. *Carnatic* music is found predominantly in the peninsular regions, and *Hindustani* music is found in the northern, eastern, and central regions.

The basic concepts of *Hindustani* classical music include *shruthi*, *textiswaras*, *alankara*, *raga*, and *tala*. In the concept of Indian music, *shruthi* means the smallest interval of pitch that the human ear can detect and a singer or musical instrument can produce. The concept of *swara* differs from the concept of *shruthi* in Indian music. A *shruthi* is the smallest gradation of the pitch while *swara* is the selected pitches from which the musician constructs the scales, melodies, and *raga*.

Alankara is any pattern of a musical piece created by the musician. It is a combination of *swara* in any creative pattern possible within an octave. While there are 7 basic notes in an octave, a slight shift in frequency for

each note gives rise to different types of existence for a single note. But the above is not true for all the seven notes. Unlike speech, music has to follow a strict range of frequencies. This range is what is called a scale. Some common types of *alankara* are *meend*, *kan-swar*, *andolan*, *gamaka*, *khatka*, and *murki*. *Alankara* is a type of exercise for *swaras* in an octave which is based on 7 main *talas* and their variation. A *raga* consists of atleast 5 notes and a *raga* can be reordered and improvised by the musician. Some of the popular *ragas* are *Bhairav*, *Yaman*, *Darbari*, and *Marwa*. The *ragas* can be subdivided into different categories, namely, morning *raga*, and evening *raga*.

Tala in Indian classical music refers to the musical beat or strike to measure the musical time. In the context of *Hindustani* classical music, *tala* is performed with clap ie. tapping one's hand on one's arm. It guides and expresses the musical rhythm and song.

Indian classical music is usually played with the two musical instruments, namely, *Harmonium* and *Tanpura*. These musical instruments are played in the background along with the vocals. The *Harmonium* provides the origin scale and *Tanpura* provides the base for the song.

Music theory is a fundamental component of human culture and behavior. That is why in many nations music training is very popular from preschool through post-secondary education. Cultures from around the world have distinctive approaches to music education. In India, institutional music education was originated by Rabindranath Tagore. Currently, there are a bunch of institutes especially dedicated to teaching music. These institutes provide music lessons in singing and multiple musical instruments like *Harmonium*, *Veena*, *Tanpura*, *Flute*, *Tabla*, and so on. Other than Indian classical music, these institutes have courses for western music too.

In Indian classical music, the concept of *shadja* or root-note is prominent. All the other musical notes in a scale are based on the *shadja* note. *Shadja* decides the base of the song and the scale in which the song has to be sung.

The knowledge of music can not only be imparted through classroom sessions but there are also many online resources available to learn music. There are several online video tutorials available on music theory and to learn different musical instruments. To learn music, one of the most important elements is feedback. Without feedback, it becomes difficult to know the learning status.

2.1.2 Vocal Melody Extraction

Melody is the essential core of music, and music without melody is unimaginable. A melody in music is a linear sequence of notes of various pitches (how high or low a note sounds) which are played one after another. The sequence of notes that comprise melody is musically satisfying and are often the most memorable part of the song. The vocal melody is defined as the sequence of notes of various pitches constructed through singing voice.

Melody extraction[6] is vital to analyze the singing skills of the user. Extracting melody from singing voice can be used for not only music retrieval, for example, query by-humming, or cover song identification but also voice separation as a guide to inform the voice source. The extraction of singing melody is a task that tracks pitch contour of the singing voice in polyphonic music[7]. Melody extraction algorithms for single-channel polyphonic music typically rely on the salience of the lead melodic instrument, considered here to be the singing voice.

Based on the assumption that vocals carry the main melody, one could apply monophonic transcription techniques to extract the melody from the separated vocal track. The majority of the algorithms are based on computing a saliency function of pitch candidates or separating the melody source from the mixture. They typically return melody as a continuous pitch stream[8].

Melodia melody extraction algorithm employs a pitch contour selection stage that relies on several heuristics for selecting the melodic output. The algorithm is comprised of four processing blocks: sinusoid extraction, saliency function computation, contour creation and characterization, and finally melody selection. In the first block, spectral peaks are detected, and precise frequencies of each component are estimated using their instantaneous frequency. In the second stage, a harmonic summation-based saliency function is computed. In the third block, the peaks of the saliency function are tracked into continuous pitch contours using auditory streaming cues. This algorithm has proven to perform particularly well on music with vocal melodies as compared to music with instrumental melodies. This indicates that the heuristic steps at the contour selection stage may be well-tuned for the singing voice, but less so for instrumental melodies.

There are other systems for melody extraction which replaces the common series of heuristic steps with a data-driven approach. Though the performance of the system nearly matches the Melodia's overall performance, this kind of system is more focussed on extracting melody from instrumental

music. Other methods to extract singing melody in polyphonic music includes harmonic tracking, salience function and so on. The harmonic tracking method is based on the subharmonic summation spectrum and harmonic structure tracking strategy[9]. However, the tracking and verification rules used in this method are quite primary which can be improved further.

One of the approaches to extract melody on vocal segments is based on classification using multi-column deep neural networks[10]. The limitation of this model is that it works well only for singing voice because it is trained with songs where vocals lead the melody. Also, since a simple energy-based singing voice detector has been used in this model, the performance of this model has limitations as compared to the previous state-of-the-art model.

Some of the previous research explores the use of source-filter models[11] for pitch estimation and their combination with voice estimation methods for automatic melody extraction. The statistics methods like Gaussian Scaled Mixture Model (GSMM) and an extended Instantaneous Mixture Model (IMM)[12] have been implemented for melody extraction.

2.1.3 Tonic Identification

Tonic in Indian classical music refers to a definite pitch value. It is the base pitch chosen by the artist to compose melodies during music composition and all the instruments are tuned using this tonic pitch[13]. The root note, “*Sa*” is known as a tonic in Indian classical music form. When considering both male and female vocalists, the frequency range in which the tonic pitch may reside spans more than one octave, typically between 100 - 260 Hz. The pitch is associated with the periodicity of the sound. It allows the arranging of sounds ranked low to high on a musical scale.

The identification of tonic pitch is the first and foremost important step to analyze any musical piece. The tonic pitch defines the scale and also the location of all other notes present in that scale. There are various approaches to identify the tonic pitch[14][15] present in audio namely, short-time Fourier transform, time domain autocorrelation function[16], harmonic source separation technique and so on.

One of the previous approaches for tonic pitch identification[17] from vocal recordings is based on the detection of spectral harmonics using predominant-F0 extraction algorithm[18]. This approach has been fairly restricted in terms that it is necessary to adapt the window length to the signal characteristics for better pitch detection accuracy. In some cases, tonic pitch detection is

done using the method based on a multi-pitch analysis[19] of the audio.

To the best of our knowledge, approaches that combine multi-pitch analysis with machine learning provide the best performance for tonic identification in most cases. In this approach, both fundamental frequency and multi-pitch salience feature are used. For extraction of the fundamental frequency, short-time Fourier transform is preferred mostly. The multi-pitch salience feature is used in order to exploit the tonal information provided. Though the approach gives more accurate results for tonic pitch identification of instruments, the correct octave is fairly unambiguous for vocal performances.

The above-mentioned approaches are restricted to recognize the tonic pitch of the musical piece, a limitation which can not be imposed if one wishes to classify the combination of notes present in the vocal performance. To classify the sequence of notes present in a musical piece, tonic identification alone is not sufficient. There is a requirement of efficient pattern recognition algorithm which works as a classifier to identify the sequence of notes in a given musical piece.

2.1.4 Pattern Recognition

This section details a number of pattern recognition techniques that have been employed in the literature.

- **Unsupervised Algorithms:** A number of unsupervised algorithms are used for classification[20] of the given signal. The implementation of these algorithms does not require training data to train the model. In classification, initially, the prominent features are extracted and similar features are grouped together. To classify similar features of the signal, distance metrics methods are used to provide the similarity measure among the features and then clustering algorithms are implemented to group similar features in one cluster.

The most simple distance metric method is Euclidean method[21]. However, this method only works if the signal is time-invariant. Since audio signal is time-series based, *Dynamic Time Warping*[22][23] is more efficient as compared to other time-invariant based distance metric methods. For Clustering of the features, k-means[21] is the simplest and most popular algorithm. The only drawback of k-means is it requires the number of clusters to be known in advance.

- **Supervised Algorithms:** To implement supervised algorithms, training data is required. Some of the supervised algorithms include Hidden Markov Model[24][25], Support Vector Machine, K-nearest neighbors, Gaussian Mixture Models[26], Mixture of Experts and so on. In some cases, the concept of neural networks ie. self-organizing map is used to identify the pattern. In previous research work, the supervised algorithms have been implemented to extract the characteristics of singing voice like harmonicity, formants, vibrato and termolo[27].

One of the previous approaches for singing voice extraction is focussed on Bayesian nonnegative matrix factorization (NMF)[28]. However, the supervised approaches normally rely on annotated data to train the classifier. This implies performing costly manual annotations of the training data.

In previous research work, semi-supervised algorithms[29] have also been explored to recognize the musical instrument. These algorithms use the semi-supervised learning techniques which are based on Gaussian mixture model and also utilise the iterative Expectation-maximization based learning. The design of the vocal training tool does not provide training data, so unsupervised algorithm has been used in the designing.

- **Other Methods:** For identifying the pattern, other methods like step-detection, boundary detection, spikes detection, and peak detection are good for the purpose of designing this tool. These methods are simple yet serve the purpose of identifying[30] general *alankara*. The only drawback is that a lot of parameters need to be calibrated efficiently to get the desired results.

2.1.5 Tuning Method

The tuning of a musical instrument is the most fundamental task. There are several methods used for tuning in different cultures of music. The equal-tempered tuning method is the most popular and straightforward method mainly used in western music. The tuning is required to identify the notes present in a scale. The tuning method defines a mathematical relation between the notes present in a scale. This relation gives the exact location of all the notes present and also helps in calculating the range for each note present in that scale.

For the development of this tool, the equal-tempered tuning method has been used as it provides the easy computational methods for calculation of notes frequency.

2.1.6 Music Performance assessment

Assessing musical performance is an essential component of music education. When a teacher guides a student in learning music or singing, to keep an account of the student's process, he/she needs to evaluate the performance of the student depending on different parameters. These parameters depend on various musical characteristics like rhythm, scale, vibrato, pitch, melody, and so on. Depending on these characteristics, one needs to define the evaluation parameters for assessing the performance of a student.

The characteristics of musical instruments, vocal music, music genre, and so on help in defining the evaluation parameters for performance assessment. Research into music performance requires a method of quantifying differences and changes between performances. Commonly, performance assessment schemes from educational contexts have been used for evaluation purposes. The relationship between various musical characteristics also helps in defining the evaluation scheme for the performance assessment. Pitch-matching accuracy, speech fundamental frequency, speech range, age, and gender are some of the characteristics which assist in performance assessment.

In the past research, the evaluation parameters, namely, pitch accuracy, number of notes in range, the correlation between various musical characteristics, musical score, and so on, assess the performance of a student. In previous music education research, real-time displays[31] have been employed to enhance singing. Many karaoke application software displays the ideal singing scale and users are supposed to practice using that scale as a reference. These applications also present the performance scores based on the divergence from the ideal defined scale. The real-time display[32] of musical notes notifies the user regarding the ideal scale of the song and also shows the notes present in the song. It helps the user in monitoring his/her performance while practicing the song. However, these applications do not produce the exact measure of the deviation from the actual notes.

The karaoke applications[33] are usually employed for fun and entertainment. These applications also present real-time song lyrics and scale of the song. *Smule* is one of the karaoke applications which display the lyrics and frequency lines of ideal notes present in the song. It is mostly used for fun

and collaborative singing purpose.

Music performance assessment[34] is also practiced for other educational purposes like learning a foreign language through singing, detecting emotions through singing, and for medical purposes. Music therapy has been explored to develop a device for the treatment of motor impairment. These devices promote exercise for the population with motor impairment and improve motor function for instrument playing into simplified hand gestures. Self-evaluation metrics present the music performance score and promotes group collaborative therapy through extrinsic feedback. In these devices, the user can calibrate their composition by mapping any sound to the output classes. However, the movements defined for rehabilitation are not clinically approved. Also, the mapping of gestures with audio feedback does not have any psychological basis to help improve the health of the user.

Foreign language learning through singing is one of the leading research topics among music researchers. Singing songs in foreign languages to develop our natural speaking is engaging and motivating. *SLIONS*, Singing and Listening to Improve Our Natural Speaking[35], is one of the software which helps in learning a foreign language through singing. It provides multi-modal instructions like auditory, visual, and reading. In the feedback system of this software, overall score and individual lyric lines present the performance assessment of the user. It corrects pronunciation through instruction, feedback, practice, and encouragement, along with improving vocabulary by providing lyrics translations and highlighting words and phrases. However, the evaluation based on singing skills has not been implemented in this software. Though it provides information about the vocabulary score and pronunciation of the user, it does not provide any information about the vocal skills of the user in singing a particular song.

The assessment analysis and scoring of the singing voice in previous literature has also been performed by computing various performance ratings like elemental ratings, timing ratings, and expression ratings[36]. The elemental rating gives the pitch and volume of the sung musical piece. The timing rating gives the segmentation and alignment of musical notes. The expression rating is based on the expressive transcription and categorization of the song.

In past research works, the scoring of the singing voice is based on melodic similarity measures. Such measures implement different MIR techniques, such as melodic transcription or score alignment[37]. In this approach, the singing voice is compared with the reference melody, which is a professional

performance of the melody. But the original score here has been used with minor changes in the schema. For pitch intonation evaluation, the teacher criteria have been very well modeled in this research work.

2.2 Challenges

In the music education, there has been several tools and applications developed for learning musical instruments and singing. One of the singing applications for Indian classical music is “*Riyaz*”. This singing application is aimed to improve the vocal skills for learning Indian classical music. It provides many basic *swaras* to practice along with a set of *alankaras* for beginner level. It also provides a set of bollywood songs to practice at an advanced level. For evaluation purposes, it has scoring system which rates the performance of the user out of 100 percent and points out the line in which the user needs to improve. But it does not provide the exact *swara* in which the user needs to improve and by how much.

There are other various similar applications for learning western music. But all these applications provide feedback which lacks the information of exact note in which the user is supposed to improve. Also, these applications have limited options of scales thus, it does not give the user the option to choose scale according to his/her voice. This provides the opportunity to work on these shortcomings of the vocal training tool and to provide a tool with more user-friendly features with detailed feedback similar to a classroom music teacher.

Chapter 3

Design and Development

This chapter briefly describes the method used in designing the vocal training tool and the algorithms implemented to develop the first version of this tool. It focuses on the important elements in designing this tool and describes the step by step process implemented in developing this tool. Initially, the chapter focuses on the root-note identification which is the first and most important step in the designing of this tool. It introduces the algorithm used in the root-note identification and clustering techniques to extract the root note from a given piece of audio. The root-note identification section also elaborates on the signal processing fundamentals implemented to extract the melody from the given audio piece.

The second section is focused on basic *alankara* identification which describes the clustering techniques used to identify the basic *alankara* pattern in the Indian classical music form. The third section is focused on general *alankara* identification which introduces the two main data processing techniques, namely, moving average and step detection to identify the general *alankara* pattern. Finally, the last section describes a summary of the design of the Indian classical vocal training tool.

3.1 Root-note Identification

For root note detection[38], initially, PredominantPitchMelodia algorithm, which uses the MELODIA[39] algorithm to estimate the fundamental frequency[40] of the predominant melody, is used. It identifies the pitch values of 5 seconds of audio signal samples on Hanning window by implementing the short-time

Fourier transform (STFT). The resulted pitch values are the combination of frequencies of the root note, vocal variations, and noise as shown in Fig. 3.1. These values imply that a major part of frequency values are consistent and these frequencies correspond to the frequencies of the root note. These pitch values are refined further to extract the ideal root note frequency.

Figure 3.1: Pitch values of audio signal (of root note) obtained using PredominantPitchMelodia.

Implementation of PredominantPitchMelodia algorithm requires defining the parameters of audio signal processing and using short time Fourier transform technique, which involves analyzing the signal frame by frame. At a given time, a particular chunk of the frame is analyzed. One Chunk variable holds 1024 frames and each frame contains 2 samples of the audio signal (since the channel is equal to 2). This implies that one chunk variable holds 2048 samples. Thus, frame size is assigned to be equal to 2048. Hop-size denotes the number of samples in between successive frames. Hop-size should be selected so that frames overlap to produce good results. Here hop-size is set to be 25 percent of frame size thus selected as 512, which produces desirable results to detect pitch values of the audio signal. For refining these pitch values to extract single root note frequency, k-means clustering algorithm[41] has been implemented. For clustering, k-means requires the number of clusters into which the sample points need to be grouped.

As mentioned above that pitch values are combination of different frequencies which can be grouped in four groups, namely, one group for frequencies of root note, two groups for vocal variations i.e. frequencies lying in range higher than root note and frequencies lying in range below root note

Figure 3.2: k-means clustering to identify the root note.

and one group for noise frequencies as shown in the Fig. 3.2. Hence, taking mean of the entire signal would not represent the singer's intended root note or scale. Thus clustering is required and the estimated number of clusters for detecting single note is four. To extract the cluster corresponding to the frequencies of the root note, the largest cluster is selected out of 4 clusters taking into account not selecting the noise cluster. It has been assumed that among the four clusters formed, the largest cluster is the one belonging to root note frequencies or noise, as deduced from the figure, the red cluster (Cluster3) belongs to the root note and blue cluster (Cluster 1) belongs to noise. The reason for this assumption is that the major part of the audio signal will be of the user singing root note and environmental noise, whereas vocal variations would be less as compared to the root note. The mean value of each cluster is checked if it is below a threshold value, which is assumed to be 50Hz here through experimentation, then it belongs to noise cluster and it is removed while selecting the largest cluster. The mean of the root note frequencies cluster gives the ideal root note frequency which is further used to calculate the ideal range frequencies in Chapter 4.

3.2 Basic Alankara Identification

Alankara Identification is done in different phases. In the first phase, the PredominantPitchMelodia algorithm from Essentia library has been used to detect the pitch of the audio signal[42]. In the next phase, these pitch values are refined to differentiate between notes present in the audio signal and identify their order.

Figure 3.3: Pitch values of Alankara audio signal obtained using PredominantPitchMelodia.

In PredominantPitchMelodia algorithm the short time Fourier transform[43][44] parameters, namely “frame size” and “hop size” are given as input for analyzing *Alankara* audio signal and their values remain the same as given in Chapter 2 (Root identification). The resulted pitch values of the audio signal (as shown in Fig. 3.3) are grouped by implementing k-means clustering[45], which helps in extracting the clusters having the note frequencies. The number of clusters per note has already been estimated in the process of root identification. Considering that, every note has 3 clusters for root note and vocal variations and 1 cluster for noise. That same concept can be used to estimate the number of clusters for multiple notes with a slight variation. In case of multiple notes, the cluster of noise for all notes is assumed to be the same, which makes the estimated number of clusters per note as 3 and one common noise cluster for all notes.

Here, the analysis has been done for n notes which produce the estimated number of clusters as $3n + 1$ (if the noise cluster is present). To extract the clusters corresponding to the frequencies of notes, n largest clusters are

Figure 3.4: k-means clustering to identify the combination of notes present in *Alankara*.

selected out of $3n + 1$ clusters taking into account not selecting the noise cluster. In Fig. 3.4, the clusters have been shown for n equal to 8. The frequency values below 50Hz are assumed to lie in noise cluster. To avoid the noise cluster coming in n largest clusters, the mean value of cluster is checked if it is below 50Hz then it is not considered while selecting largest clusters. The concept behind selecting the n largest clusters is that clusters having notes' frequency would be largest as it has been assumed that vocal variations would be less compared to the note frequencies. The pitch values of other clusters (apart from n largest clusters) are assigned zero values to easily extract note frequency.

These n clusters have all the information of notes present in the audio signal. This information has been used to check the sequence of notes sung by the user. This has been achieved by extracting the starting index of each cluster and arranging those indices in the ascending order which is basically arranging the clusters according to their occurrence. Thus it gives the order of clusters according to the sequence of notes, as shown in Fig. 3.5. Further, the mean of these arranged clusters denotes the frequency of the note in sequential order. For this, each cluster mean has been checked whether it is lying in the range of ideal note frequency (calculated in Chapter 4) or not. Other evaluation parameters have also been calculated, discussed in section D, to help the user learn more about their vocal skills.

Figure 3.5: Clusters of n notes to obtain sequence of notes and their comparison with the ideal range of notes for calculating voice stability and pitch accuracy.

3.3 General Alankara Identification

In the previous chapter, the concept of *alankara* has been described as a combination of different musical notes. There are several possible combinations of notes corresponding to different *alankara*. Among these possible combinations, the basic *alankara* (*Sa Re Ga Ma Pa Dha Ni Sa*) has been identified by implementing PredominantPitchMelodia algorithm for frequency extraction and k-means clustering method for recognizing the sequence of notes.

The general *alankara* can have several possible combinations of notes present in an octave. These possible combinations can also have a repetition of the notes. To identify the *alankaras* owning the repetition of notes is a quite challenging task. Suppose if the vocalized *alankara* is (*Sa Re Ga, Re Ga Ma, Ga Ma Pa, Ma Pa Dha, Pa Dha Ni, Dha Ni Sa*). This *alankara* includes repetition of notes. For identifying this kind of *alankaras*, a better approach is required.

Firstly, the dominant pitch values are extracted from the audio signal using PredominantPitchMelodia algorithm. Next, the tool requires to identify the sequence of notes from these pitch values. To identify this sequence,

the k-means clustering method is not a valid option as it clusters the two different occurrences of the same note in a single cluster which makes it difficult to identify the actual sequence. Since the k-means clustering method forms clusters based on the frequency value of the note and does not take into account the time factor, there is a requirement of a better approach.

The new approach to identify the general *alankara* should be able to classify the notes based on their occurrences and correctly identify the actual sequence of the *alankara*

Figure 3.6: Moving average to smoothen the signal.

Initially, the tool obtains the pitch values of the audio signal by applying the PredominantPitchMelodia algorithm of essentia library. The identification of shift from one note to another is the challenging task in recognizing the *alankara*. To identify this shift, the tool must detect the sudden transitions present in the signal. The position of abrupt transformations represents the boundary of notes and give information about the duration of each note. To extract the location of sharp change, a set of methods have been implemented.

Firstly, the obtained set of pitch values are normalized using the moving average method. Next, the step-detection algorithm has been implemented to extract the *alankara* sequence. The following section elaborates the above-mentioned method in detail.

3.3.1 Moving Average

The moving average is the arithmetic mean of the n number of points. The moving average modifies the set of pitch values extracted using PredominantPitchMelodia algorithm. The audio signal becomes smoother after the implementation of the moving average method, which makes it easier to detect the position of abrupt transformations by further implementing the step-detection algorithm.

The resulted number of pitch values is less than the original number of pitch values. This decreased number of pitch values is stored as the margin and this margin is summed to the obtained indices through the step-detection algorithm so as to obtain the exact position of the note change.

The moving average makes the change of notes more obvious and accessible. Since the entire audio signal smoothens, the duration for which the user is singing a particular note grows constant and the position where the note is varying becomes like a gradient.

Figure 3.7: Moving average to smoothen the signal.

Here the analysis has been done for the general alankara having 18 notes in total (including the repeated notes). The Fig. 3.7 shows the moving average of obtained original pitch values from the audio signal. Here the shift from one note to another is apparent as compared to the Fig. 3.6, which depicts the original pitch values before implementation of the moving average. Also computation-wise, it is easier to identify the sequence of notes after the implementation of moving average.

After analyzing several samples, the parameter n has been calibrated to

be equal to 15. This optimal value of n gives the desired result for most of the audio samples.

3.3.2 Step-detection

By executing the step-detection method, the tool identifies the indices where the shift from one note to another is occurring. Since the indices of note variation can be extracted through the gradient present at the position of note change, the step-detection algorithm has been used for this purpose.

In step-detection algorithm, initially, the one-dimensional Gaussian filter is employed. The standard deviation value of the Gaussian kernel is calibrated to be equal to 6 to achieve desired results. The order of the Gaussian filter is set to be equal to one according to the specification. After the implementation of the Gaussian filter, the steps (note variation) becomes prominent in the resulted signal.

Figure 3.8: Implementation of Gaussian Filter to identify the steps.

Further, these noticeable steps are obtained to recognize the sequence of notes present in *alankara*. To extract these obvious steps, a threshold value has been defined. The absolute value of the resulted pitch values after the implementation of the moving average gives the threshold value. The pitch values traversing this threshold depicts the prominent peaks.

In this manner, the indices of steps have been obtained. Among these indices, there is a possibility of some indices which does not actually represent

note variation. To remove these extra indices, the step size of each step has been calculated and the extra steps with the least step size are removed from the obtained list of steps. This gives the exact number of note variation present in the *alankara*.

Figure 3.9: Implementation of Threshold to identify the prominent steps.

In Fig. 3.9, the analysis has been done for the general *alankara* detection having a combination of 18 notes. The cross sign depicts the presence of note variation in the audio sample. The location of these variations are accurately identified by the implementation of above-mentioned method. This experiment has been repeated for different set of samples to calibrate different parameters involved in the implementation of step detection algorithm.

These indices of note variation are used to extract the information about each note in *alankara* and compute the evaluation parameters ie. pitch accuracy and voice stability.

3.4 Summary

The vocal training tool for Indian classical music has been designed in a user-friendly manner. It adjusts the scale according to the vocal scale of the user. In the first stage of the development of this tool, the root-note identification has been done. For root-note identification, initially signal processing techniques are implemented and then the PredominantPitchMelodia algorithm has been implemented which filters out the background noise and extracts the melody from the musical piece. This melody is further analyzed using

the k-means clustering technique to extract the root-note of the vocal voice of the user. Similar approach has been used for basic *alankara* identification. Though this approach provided accurate results for basic *alankara* identification, it does not work for the general *alankara* identification. Thus, for general *alankara* identification, moving average and step detection algorithm has been used.

Chapter 4

Results and Discussion

This chapter briefly describes the evaluation methods used for music performance assessment and discusses the results obtained from different experiments conducted. It describes the parameters, namely, notes in range, voice stability, and pitch accuracy, in detail. Then it focusses on experimental results obtained from different subjects. The experiment section is subdivided into two subsections, namely, basic *alankara*, and general *alankara*. The basic *alankara* section discusses the results obtained for experiments on audio recordings having basic *alankara* that is “Sa Re Ga Ma Pa Dha Ni SA”. The general *alankara* section discusses the experiments conducted on audio recording having more complicated *alankara* pattern. The pattern used for general *alankara* in this research is “SaReGa ReGaMa GaMaPa MaPaDha PaDhaNi DhaNiSA”. The chapter ends with a summary summing-up the results obtained from the experiments conducted.

4.1 Ideal Range Calculation

The root note identified in Chapter 2 is assigned as the ideal root note, which is used to calculate the other reference notes through a mathematical relation [46] followed by musical notes. These notes are termed as reference notes because they are used to compare the detected note frequencies in Chapter 3.

The mathematical relation between frequencies of notes of equal-tempered scale is given by:

- $f_n = f_0 * a^n$ where,

- f_0 = the ideal root note frequency.
- n = the number of half steps away from the fixed note. If you are at a higher note, n is positive. If you are on a lower note, n is negative.
- f_n = the frequency of the note n half steps away.
- $a = 2^{\frac{1}{12}}$

Using this relation, the note frequency for each note is calculated and set as a reference for comparison. Due to large variations in vocal frequencies, the identified note frequencies might not be exactly the same as reference note frequency, this brings the concept of calculating the range for each note. Every note has a particular range of frequencies lying in that range. The frequencies lying outside this range does not belong to that note. This range has been calculated by measuring half of the difference between two consecutive notes and it would vary for different notes because the difference between two notes varies.

Suppose the note frequency of root note is x and the frequency of semi-note before and after x , is y and z respectively then the range of x is defined as: $[(x - y)/2, (z - x)/2]$. Here $y = x \times a^{-1}$ and $z = x \times a$. This range calculation gives an ideal range for each note which is used to check whether the identified note lies in this range or not. It also helps in comparison for the deviation of identified note, from the ideal, which provides voice stability and pitch accuracy parameters for evaluation.

4.2 Evaluation Criteria

The following parameters have been defined for the end user in an understandable form which gives the knowledge to the user on how to improve their vocal skills:

4.2.1 Notes in Range

This evaluation parameter gives the number of notes user is singing in the defined scale. This parameter tells the exact note in which the user is going out of scale. It provides the user an overview of the mistakes he has done while singing a particular combination of notes.

4.2.2 Voice Stability

This evaluation criterion gives the users a clear picture of the stability in their voice. More the voice is stable, better is vocal stability of the person. To sing a particular song melodiously, the first and foremost step for an individual is to bring stability in their voice. The voice stability depends on the number of noise points within that time frame.

For a particular note, the group of values identified in *Alankara* identification is compared with the ideal range of that note calculated in Chapter 4. This gives the number of frequencies lying in the range and this process is repeated for each note. The percentage of the total frequency points lying in the ideal range out of total data points gives voice stability of the user.

Suppose the detected combination is Sa Re Ga Ma Pa Dha Ni SA. From this the length of the identified group of values for Sa note is n . Now each of these n values is compared with the ideal range of Sa and among n values, n_1 are lying in the defined range. Similarly, for Re note, m_1 are lying in Re range out of m values. This process is repeated for each note. Now to calculate voice stability, we will sum up these values to check how many values are lying in their defined range.

4.2.3 Pitch Accuracy

The pitch accuracy provides the exact measure of the deviation from the ideal pitch of the note. This parameter helps the user to know how much flat or sharp the note has to go to match the ideal note. Pitch accuracy depends on the overall closeness of the non-noisy points to the ideal note frequency. So to calculate this parameter, the tool finds the difference between the ideal note frequency and identified note frequency of the user. If the calculated pitch accuracy is zero then it is ideal.

If the value of calculated pitch accuracy is equal to or higher than one semitone difference, then it means the user has deviated to sharp note. Based on this pitch accuracy the tool will predict the note on which the user has deviated. Similarly, if the pitch accuracy is negative and its magnitude is equal to or higher than one semitone then it means the user has deviated to flat note.

This parameter is really helpful while practicing a song. Through pitch accuracy, the user can know that in which note of the song he is going flat or sharp and by how many semi notes.

4.3 Experimental Results

4.3.1 Basic Alankara

The proposed method of identifying the basic alankara and presenting the analysis of the user's voice in terms of the notes lying in the ideal range, voice stability, and the pitch accuracy has been implemented and the experiment has been repeated for 4 different sets of vocal recordings and 1 musical instrument recording as shown in Table 4.1, 4.2, and 4.3. In Table 4.2, and 4.7, the term “True” means that the sung note is lying in the range of ideal note and “False” depicts that the sung note is lying outside the range of ideal note.

The result shows that all the notes of *Shraddha1*, *Sagar* and *Harmonium* recordings are lying in the ideal range, which means that these three recordings have all the notes within the defined scale. The pitch accuracy of these three recordings is also observed to be close to zero, which means that the detected note frequency value of these recordings is quite close to the ideal frequency of the note. The positive sign of the pitch accuracy tells that the user is singing higher than the ideal note frequency by the specified value and negative sign means that the user is singing lower than the ideal note by a given value.

As the pitch accuracy is a measure of the closeness of the detected frequency with the ideal frequency, so for the subject *John*, the magnitude of pitch accuracy is very high for every note. Higher the pitch accuracy, farther the note from the ideal note. This is the reason that for *John*, all the notes are outside the ideal range except the first note. Since the first note defines the scale in which the user is singing, this means that all the other notes are not on the scale defined by the first note for *John*. The result of the subject *John* can be used as the identifier for different medical applications of this tool.

Shraddha2 recording has all the notes in the ideal range except the 4th note, which is visible in the pitch accuracy of the 4th note (*Ma*). For the 4th note, the magnitude of pitch accuracy is the highest as compared to other notes of the subject *Shraddha2* recording which again tells that detected note frequency of the 4th note is quite away from the ideal frequency of the 4th note. For “Ma” note of *Shraddha2* recording, the user is singing 15.144 Hz higher than the ideal note frequency, which is quite high as compared to other notes.

Table 4.1: Results showing Voice stability and number of notes in ideal range for 5 subjects of basic alankara

Dataset	Notes detected in ideal range	Voice Stability(%)
Shraddha1	8/8	57.070
Shraddha2	7/8 (except 4th note)	49.275
Sagar	8/8	59.438
John	1/8	21.372
Harmonium	8/8	82.101

Table 4.2: Notes in range for 5 subjects of basic alankara

S.No.	Shraddha1	Shraddha2	Sagar	John	Harmonium
Sa	True	True	True	True	True
Re	True	True	True	False	True
Ga	True	True	True	False	True
Ma	True	False	True	False	True
Pa	True	True	True	False	True
Dha	True	True	True	False	True
Ni	True	True	True	False	True
SA	True	True	True	False	True

The pitch accuracy values for *Harmonium* recording is best in terms that the detected note frequencies are almost equal to the ideal frequency of notes. The *Harmonium* recording is the instrumental recording which is having the fixed note frequencies and vocal variations are not present in instrument recordings thus it is giving the good pitch accuracy. It means that the tool is correctly identifying and comparing the notes. Also, the voice stability is highest for *Harmonium* recording, which is again because of the fixed frequency of the notes in the instrument.

Table 4.3: Pitch Accuracy of 5 subjects in Hz for basic alankara

S.No.	Shraddha1	Shraddha2	Sagar	John	Harmonium
Sa	0.0	0.0	0.0	0.0	0.0
Re	-4.523	-0.627	-2.078	-261.628	-0.0002
Ga	3.367	1.728	-1.651	-293.667	-0.0004
Ma	6.387	15.144	0.312	-329.631	-0.0006
Pa	-0.148	-1.153	-4.118	-349.232	-0.001
Dha	5.865	3.453	-0.997	-391.999	-0.001
Ni	8.674	3.307	-4.362	-440.005	-0.002
SA	7.032	11.191	-0.417	-493.889	-0.002

4.3.2 General alankara

The proposed method of identifying the general alankara and presenting the analysis of the user's voice in terms of the notes lying in the ideal range, voice stability, and the pitch accuracy has been implemented and the experiment has been repeated for different sets of vocal recordings as shown. The analysis has been done for one particular general *alankara* at this stage and this *alankara* comprises repetition of *Shudha swaras*. The number of notes in this general *alankara* (including the repetition of notes) is 18.

The results in Table 4.4 shows that for the audio recordings of *Sagar* and *Mahadevan*, all the 18 notes are lying in the ideal range. The voice stability of *Sagar*, Mahadevan and *Shraddha2* is quite good and above 80 percent. This means that these users have good vocal stability and their voice is more stable as compared to other users. For *Shraddha2* recording, even though four notes are lying outside the range, the voice is more stable and it shows that this user already has the voice stability and needs to work on singing

the notes in the ideal range sustaining this stability in the voice.

Table 4.4: Results showing Voice stability and number of notes in ideal range for 5 subjects of general alankara

Dataset	Notes detected in ideal range	Voice Stability(%)
Sagar	18/18	82.343
Mahadevan	18/18	81.314
Akhilesh	8/18	33.403
Shraddha1	15/18	77.057
Shraddha2	16/18	81.597

For the audio recording of *Shraddha1*, three notes are lying outside the ideal range. The exact notes lying outside the range for this user are 2, 4 and 16. And the voice stability for this recording is 77.057 percent. Analyzing the audio recordings of *Shraddha1* and *Shraddha2*, it has been observed that voice stability is not directly proportional to the number of notes lying in the ideal range. The voice stability for the recording *Shraddha1* is lower as compared to the recording *Shraddha2* even though the number of notes lying in the ideal range is greater for the former.

The result for the audio recording *Akhilesh* shows that the number of notes lying in the ideal range and the voice stability, both are less for this user. This means that the user needs to work on both singing the note correctly and bringing stability in the voice.

Table 4.5: Results showing Voice stability and number of notes in ideal range for instrumental recordings of general alankara

Dataset	Notes detected in ideal range	Voice Stability(%)
Harmonium1	18/18	99.37
Harmonium2	18/18	98.67
Flute1	18/18	99.90
Flute2	17/18	95.20

In Table 4.5, Harmonium1, and Harmonium2 represent the recordings from the harmonium instrument in two different scales. The harmonium recordings have been taken from the Yamaha PSR e353. The pitch pattern in the harmonium instrument is similar to the vocal voice of human and thus it is used for teaching Indian classical music. To verify and validate

the design of this tool, two different harmonium recordings have been tested. The scale of both recordings is different. The first recording is having a fundamental note of frequency 276.753 Hz and the second recording is having a fundamental note of frequency 261.233 Hz.

In Table 4.5, Flute1, and Flute2 recording represent the audio recording of the flute, which is an Indian classical instrument. The recording has been used to validate the design of this tool similar to the harmonium recording. As expected, for the Flute1 recording, all the notes are found to be in the ideal range and for Flute2 recording, one note is found to be lying outside the selected ideal scale. In the second recording of the flute, the 5th note was deliberately altered to verify the design of this tool and as expected, the tool points out the wrong note while calculating the results. As observed in Table 4.5, all the notes are in range for Harmonium1, Harmonium2, and Flute1. Also as expected, voice stability is more than 95 percent which is quite accurate. This goes in line with the fact that the harmonium and flute instrument is good for learning Indian classical music as it improves voice stability and also improves the pitch accuracy. As expected, the voice stability of the instruments is found to be close to 100 percent which verifies and strengthens the design of this vocal training tool.

As shown in Table 4.6, the pitch accuracy of notes is diversifying for the recording *Akhilesh*. For the notes lying in the ideal range, the pitch accuracy is close to zero. While for the notes lying away from the ideal range, the magnitude of pitch accuracy raises. It means that magnitude depicts the factor by which the sung note has deviated from the ideal value. This parameter is showing the exact measure of deviation from the ideal note. The pitch accuracy is directly proportional to the number of notes lying in the range. For the audio recording of *Sagar*, the least variations are present in the pitch accuracy and the magnitude of the pitch accuracy is below six for each note. Thus all the notes are lying on the defined scale for this user. As it has been observed from Table 4.7, for *Shraddha1* recording, the 2, 4 and 16 note is lying outside the range thus the magnitude of pitch accuracy for these notes is quite high. But there are notes with a higher magnitude of pitch accuracy and lying in the ideal range. This is because the ideal range for each note increases while moving towards a succeeding note. That is, the range of higher frequency note is also large as compared to the range of lower frequency note.

From Table 4.8, it has been observed that pitch accuracy for all the notes is close to zero. This is in line with the previous results of notes in range,

Table 4.6: Pitch Accuracy of 5 subjects in Hz for general alankara

S.No.	Sagar	Mahadevan	Akhilesh	Shraddha1	Shraddha2
Sa	0.0	0.0	0.0	0.0	0.0
Re	-3.636	-3.800	-9.209	-7.001	1.477
Ga	-0.600	-3.572	-11.195	0.143	-0.805
Re	-3.831	-2.281	-15.436	-19.246	-7.956
Ga	0.449	-2.462	-9.398	2.405	-3.666
Ma	0.834	-1.677	-10.375	4.471	-1.245
Ga	2.013	-1.885	-2.629	-2.916	-4.079
Ma	-1.729	-1.563	-10.926	-0.409	-4.911
Pa	-2.938	-4.742	-12.267	-0.304	-6.693
Ma	-3.860	-5.016	3.858	-1.689	-9.890
Pa	-2.574	-4.469	-6.636	0.949	-2.653
Dha	-1.542	-5.215	-7.349	-3.995	-5.225
Pa	-1.018	-4.505	-9.754	-8.902	-0.101
Dha	-2.367	-4.403	-6.557	-8.389	-5.314
Ni	-5.493	-4.453	-3.869	-4.005	-1.660
Dha	0.252	-6.854	-5.485	-28.065	-18.743
Ni	-3.366	-3.566	0.074	-8.301	-0.775
SA	-4.259	-6.597	-3.249	-9.079	1.330

Table 4.7: Notes in range for 5 subjects of general alankara

S.No.	Sagar	Mahadevan	Akhilesh	Shraddha1	Shraddha2
Sa	True	True	True	True	True
Re	True	True	False	False	True
Ga	True	True	False	True	True
Re	True	True	False	False	False
Ga	True	True	False	True	True
Ma	True	True	False	True	True
Ga	True	True	True	True	True
Ma	True	True	False	True	True
Pa	True	True	False	True	True
Ma	True	True	True	True	False
Pa	True	True	False	True	True
Dha	True	True	False	True	True
Pa	True	True	False	True	True
Dha	True	True	True	True	True
Ni	True	True	True	True	True
Dha	True	True	True	False	True
Ni	True	True	True	True	True
SA	True	True	True	True	True

Table 4.8: Pitch Accuracy of two different scales of Harmonium in Hz

S.No.	Harmonium1	Harmonium2
Sa	0.0	0.0
Re	0.178	0.205
Ga	-0.230	-0.102
Re	-6.097	-0.311
Ga	-0.545	2.632
Ma	-0.280	-0.248
Ga	-0.583	-0.847
Ma	0.179	-0.194
Pa	-0.270	-0.060
Ma	-1.119	-0.978
Pa	0.159	0.196
Dha	-0.209	-0.141
Pa	-1.535	-1.715
Dha	0.1840	0.333
Ni	-0.733	-0.104
Dha	-1.547	-11.154
Ni	-0.209	-0.682
SA	-0.749	-0.421

where all the notes for Harmonium1 and Harmonium2 are in range. Since the magnitude of pitch accuracy results is close to zero, thus the voice stability is also greater for these two recordings. As observed from Table 4.9, the pitch accuracy of all the notes for Flute1 is close to zero while the pitch accuracy of the 5th note of Flute2 is -18.617, which is quite far from zero. This means that the 5th note of Flute2 has been played 18.617 Hz lower than the ideal note frequency. Since this magnitude is quite high compared to the pitch accuracy of other notes, it is clear that the 5th note of the Flute2 is lying out of the defined ideal range.

Table 4.9: Pitch Accuracy of two different scales of Flute in Hz

S.No.	Flute1	Flute2
Sa	0.0	0.0
Re	0.692	0.125
Ga	0.729	0.318
Re	-0.207	-0.586
Ga	1.094	-18.617
Ma	0.797	0.103
Ga	0.101	-1.890
Ma	0.914	0.426
Pa	0.949	0.476
Ma	-0.615	-0.544
Pa	1.343	0.663
Dha	0.961	0.333
Pa	-0.242	-0.746
Dha	1.329	0.739
Ni	1.159	0.365
Dha	-0.402	-0.797
Ni	1.458	0.779
SA	1.173	-0.110

4.4 Summary

The evaluation parameters give the accurate results as expected in an understandable form for the user. It fulfils the goal of this tool by pointing out

the exact note in which the user is singing out of defined ideal scale and also gives the exact amount by which he/she needs to improve in terms of pitch accuracy. The pitch accuracy is very helpful as it compares the note user is singing with the ideal one and how much it has deviated from the ideal note. Voice stability gives the user an overall idea of the fluctuations present in the voice while singing. It helps in enhancing the voice and making it more melodious. The negative sign of pitch accuracy depicts that the user is singing at a lower frequency than the defined ideal note frequency. The lower frequency means the user is going flat than the actual note. While the positive sign means the user is singing at a higher frequency than the defined ideal note frequency. The higher frequency means the user is singing sharp note than the actual note.

Chapter 5

Conclusions and Future Extensions

This chapter provides a conclusion obtained from this research work. It elaborates the strengths and shortcomings of the Indian classical vocal training tool. In the first section, it compares the two different methods implemented in the designing of this tool and focusses on the results obtained from both the methods. It also mentions the evaluation parameters defined for the design of this tool. The second section focusses on the future work and the extension of the vocal training tool to make it more efficient. It discusses the similar research possible in other styles of music like western music, Chinese classical music, and so on. In the last section, various applications of this tool has been mentioned in the field of entertainment, medical, music education, and so on.

5.1 Conclusions

In this research work, the PredominantPitchMelodia algorithm has been employed to identify the pitch values of the audio signal. These pitch values are further grouped in order to identify the *alankara* present in the audio signal. The identified note frequencies of the *alankara* are analyzed further to calculate the evaluation parameters like voice stability and pitch values. These parameters give knowledge about the user's vocal skills in understandable form and thus help the users in learning the Indian Classical Music. As the result shows, the pitch accuracy provides the exact note in which the user

is going out of scale and also the magnitude by which the user is deviating from the ideal note. Voice stability is also one of the important parameters for evaluation as it depicts the fluctuations present in the user's voice while singing a particular piece of music.

It has been concluded that the k-means clustering method is suitable to identify the basic *alankara*. The general *alankara* can have a repetition of the notes which leads to the failure of the k-means clustering method. To identify the general combination of notes, moving average and step-detection algorithm proves to be a better approach. The implementation of moving average requires the calibration of parameters, which has been calculated by observation of a different set of samples. The step-detection algorithm also requires the calibration of parameters to detect the prominent peaks representing the indices of variation from one note to another.

For detecting the general *alankara*, a Gaussian kernel filter has been implemented to detect the variation present from one note to another. The Gaussian kernel parameters have also been calibrated to provide desired results. The harmonium instrument frequency has been set as ground truth as the frequency of instrument remains fixed while vocal frequency is varying. The tool identifies the pattern of the notes correctly and also provides an accuracy of more than 98 percent for the harmonium instrument. The results obtained for different recordings are found to be in line with the expected outcome. For general *alankara* identification, the different parameters need to be calibrated which makes the design quite complex.

The evaluation parameters defined for the design of the Indian classical vocal training tool, gives the user complete information of their singing skills in terms of stability and accuracy. It also points out the exact note in which the user needs to improve. The evaluation criteria, namely, notes in range, voice stability, and pitch accuracy, have been defined to provide the information to the user in an understandable form.

5.2 Future Work

Currently, the equal temperament tuning method has been used to calculate the ideal range. For future work, Just Intonation tuning method can be used as it represents Indian classical music more accurately. For time-series clustering, better algorithms can be used which will make a significant difference in designing the tool. The current algorithm of k-means is not suitable for

time series signal. The future algorithm should be time series based like *Dynamic Time Warping* [23] [47] to add some tolerance in time. And the time series based clustering algorithm could be combined with some adaptive window to increase the accuracy of the pattern recognition. This adaptive window could depend on some features extracted from the signal after segmentation of the audio signal. Also, for detecting the prominent steps, a better step-detection algorithm can be used as this algorithm requires the calibration of parameters.

In the future, the tool can also be extended to have additional features of converting a given piece of music or song into its basic formation in Indian classical music. The basic formation will define the scale and rhythm of the song which will expedite the singing lesson for the user. Also, the additional evaluation parameters can be used to provide a detailed analysis of the user's vocal ability. The current evaluation parameters can also be modified further to provide information in a more understandable form. For example, the pitch accuracy can be modified to notify the user that the user has deviated to which note. So, this evaluation parameter can be modified to give information in terms of musical notes rather than in mathematical form. This will help the user to understand that how much flat or sharp he has deviated from the ideal defined scale.

5.3 Applications

This tool can serve the purpose of assisting people learning Indian classical music. Also it can be used for other musical forms like western music, Chinese music and folk music. The approach to learn any music form is almost similar, only the music notation is different. So it can be used for any music form. There are various applications in which this tool can be of help. In music education applications, this tool can be used for learning a foreign language through singing. This not only helps in learning a new language but also improves the singing skills of the user by focussing on vocal skills like vocal stability and pitch accuracy of the user while he/she is singing the foreign song.

Later, this tool can be developed for medical purposes like to cure depression, neural disorders (eg. Schizophrenia) and for meditation as well. To cure motor impairment, music therapy is also used. This tool will be helpful in medical applications related to music therapy. For example, yoga and

meditation use vocal singing to cure various diseases including depression, anxiety, and other various health issues.

Author's Publication(s)

- **S. Sharma**, and M. J. Er, "Design and Development of an Indian Classical Vocal Training Tool," *2019 2nd International Conference on Intelligent Autonomous Systems (ICoIAS)*, Singapore, Singapore, 2019, pp. 58-62. doi: 10.1109/ICoIAS.2019.00017.

Bibliography

- [1] D. Bogdanov, N. Wack, E. Gómez Gutiérrez, S. Gulati, P. Herrera Boyer, O. Mayor, G. Roma Trepas, J. Salamon, J. R. Zapata González, and X. Serra, “Essentia: An audio analysis library for music information retrieval,” in *Britto A, Gouyon F, Dixon S, editors. 14th Conference of the International Society for Music Information Retrieval (ISMIR); 2013 Nov 4-8; Curitiba, Brazil.[place unknown]: ISMIR; 2013. p. 493-8.* International Society for Music Information Retrieval (ISMIR), 2013.
- [2] C. Dittmar, E. Cano, J. Abeßer, and S. Grollmisch, “Music Information Retrieval Meets Music Education,” in *Multimodal Music Processing*, ser. Dagstuhl Follow-Ups, M. Müller, M. Goto, and M. Schedl, Eds. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2012, vol. 3, pp. 95–120. [Online]. Available: <http://drops.dagstuhl.de/opus/volltexte/2012/3468>
- [3] S. Giraldo, G. Waddell, I. Nou, A. Ortega, O. Mayor, A. Perez, A. Williamson, and R. Ramirez, “Automatic assessment of tone quality in violin music performance,” *Front. Psychol.* 10: 334. doi: 10.3389/fpsyg, 2019.
- [4] J. Abeßer, J. Hasselhorn, C. Dittmar, A. Lehmann, and S. Grollmisch, “Automatic quality assessment of vocal and instrumental performances of ninth-grade and tenth-grade pupils,” in *Proceedings of the International Symposium on Computer Music Multidisciplinary Research (CMMR)*, 2013, pp. 975–988.
- [5] K. Pati, S. Gururani, and A. Lerch, “Assessment of student music performances using deep neural networks,” *Applied Sciences*, vol. 8, no. 4, p. 507, 2018.

- [6] V. Rao and P. Rao, “Vocal melody extraction in the presence of pitched accompaniment in polyphonic music,” *IEEE transactions on audio, speech, and language processing*, vol. 18, no. 8, pp. 2145–2154, 2010.
- [7] S. Vembu and S. Baumann, “Separation of vocals from polyphonic audio recordings.” in *ISMIR*. Citeseer, 2005, pp. 337–344.
- [8] R. M. Bittner, J. Salamon, S. Essid, and J. P. Bello, “Melody extraction by contour classification.” in *ISMIR*, 2015, pp. 500–506.
- [9] C. Cao, M. Li, J. Liu, and Y. Yan, “Singing melody extraction in polyphonic music by harmonic tracking.” in *ISMIR*, 2007, pp. 373–374.
- [10] S. Kum, C. Oh, and J. Nam, “Melody extraction on vocal segments using multi-column deep neural networks.” in *ISMIR*, 2016, pp. 819–825.
- [11] J. J. Bosch, R. M. Bittner, J. Salamon, and E. Gómez Gutiérrez, “A comparison of melody extraction methods based on source-filter modelling,” in *17th International Society for Music Information Retrieval Conference; 2016 Aug 7-11; New York, United States*. International Society for Music Information Retrieval (ISMIR), 2016.
- [12] J.-L. Durrieu, G. Richard, B. David, and C. Févotte, “Source/filter model for unsupervised main melody extraction from polyphonic audio signals,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 564–575, 2010.
- [13] S. Gulati, J. Serra, and X. Serra, “An evaluation of methodologies for melodic similarity in audio recordings of indian art music,” in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 678–682.
- [14] A. Bellur, V. Ishwar, X. Serra, and H. A. Murthy, “A knowledge based signal processing approach to tonic identification in indian classical music,” in *Serra X, Rao P, Murthy H, Bozkurt B, editors. Proceedings of the 2nd CompMusic Workshop; 2012 Jul 12-13; Istanbul, Turkey. Barcelona: Universitat Pompeu Fabra; 2012. p. 113-118*. Universitat Pompeu Fabra, 2012.
- [15] G. K. Koduri, J. Serrà Julià, and X. Serra, “Characterization of intonation in carnatic music by parametrizing pitch histograms,” in *Gouyon F*,

Herrera P, Martins LG, Müller M. *ISMIR 2012: Proceedings of the 13th International Society for Music Information Retrieval Conference; 2012 Oct 8-12; Porto, Portugal. Porto: FEUP Edições, 2012.* International Society for Music Information Retrieval (ISMIR), 2012.

- [16] M. A. Raju, B. Sundaram, and P. Rao, “Tansen: a query-by-humming based music retrieval system,” in *Proceedings of the national conference on communications (NCC)*, 2003.
- [17] S. Gulati, J. Salamon, and X. Serra, “A two-stage approach for tonic identification in indian art music,” in *Proceedings of the 2nd CompMusic Workshop; 2012 Jul 12-13; Istanbul, Turkey. Barcelona: Universitat Pompeu Fabra; 2012. p. 119-127.* Universitat Pompeu Fabra, 2012.
- [18] J. C. Ross, T. Vinutha, and P. Rao, “Detecting melodic motifs from audio for hindustani classical music.” in *ISMIR*, 2012, pp. 193–198.
- [19] J. Salamon, S. Gulati, and X. Serra, “A multipitch approach to tonic identification in indian classical music,” in *Gouyon F, Herrera P, Martins LG, Müller M. ISMIR 2012: Proceedings of the 13th International Society for Music Information Retrieval Conference; 2012 Oct 8-12; Porto, Portugal. Porto: FEUP Edições; 2012.* International Society for Music Information Retrieval (ISMIR), 2012.
- [20] L. Lu, H. Jiang, and H. Zhang, “A robust audio classification and segmentation method,” in *Proceedings of the ninth ACM international conference on Multimedia.* ACM, 2001, pp. 203–211.
- [21] A. Singh, A. Yadav, and A. Rana, “K-means with three different distance metrics,” *International Journal of Computer Applications*, vol. 67, no. 10, 2013.
- [22] E. Keogh and C. A. Ratanamahatana, “Exact indexing of dynamic time warping,” *Knowledge and information systems*, vol. 7, no. 3, pp. 358–386, 2005.
- [23] M. Müller, “Dynamic time warping,” *Information retrieval for music and motion*, pp. 69–84, 2007.

- [24] W. Chai and B. Vercoe, “Folk music classification using hidden markov models,” in *Proceedings of international conference on artificial intelligence*, vol. 6, no. 6.4. sn, 2001.
- [25] Y. Qi, J. W. Paisley, and L. Carin, “Music analysis using hidden markov mixture models,” *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5209–5224, 2007.
- [26] J.-C. Wang, Y.-H. Yang, H.-M. Wang, and S.-K. Jeng, “Modeling the affective content of music with a gaussian mixture model,” *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 56–68, 2015.
- [27] L. Regnier and G. Peeters, “Singing voice detection in music tracks using direct voice vibrato detection,” in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 1685–1688.
- [28] P.-K. Yang, C.-C. Hsu, and J.-T. Chien, “Bayesian singing-voice separation,” in *ISMIR*, 2014, pp. 507–512.
- [29] A. Diment, T. Heittola, and T. Virtanen, “Semi-supervised learning for musical instrument recognition,” in *21st European Signal Processing Conference (EUSIPCO 2013)*. IEEE, 2013, pp. 1–5.
- [30] C. Panagiotakis and G. Tziritas, “A speech/music discriminator based on rms and zero-crossings,” *IEEE Transactions on multimedia*, vol. 7, no. 1, pp. 155–166, 2005.
- [31] D. M. Howard, J. Brereton, G. F. Welch, E. Himonides, M. DeCosta, J. Williams, and A. W. Howard, “Are real-time displays of benefit in the singing studio? an exploratory study,” *Journal of Voice*, vol. 21, no. 1, pp. 20–34, 2007.
- [32] G. F. Welch, D. M. Howard, E. Himonides, and J. Brereton, “Real-time feedback in the singing studio: an innovatory action-research project using new voice technology,” *Music Education Research*, vol. 7, no. 2, pp. 225–249, 2005.
- [33] W.-H. Tsai and H.-C. Lee, “Automatic evaluation of karaoke singing based on pitch, volume, and rhythm features,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1233–1243, 2011.

- [34] A. Vidwans, S. Gururani, C.-W. Wu, V. Subramanian, R. V. Swaminathan, and A. Lerch, “Objective descriptors for the assessment of student music performances,” in *Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio*. Audio Engineering Society, 2017.
- [35] D. Murad, R. Wang, D. Turnbull, and Y. Wang, “Slions: A karaoke application to enhance foreign language learning,” in *2018 ACM Multimedia Conference on Multimedia Conference*. ACM, 2018, pp. 1679–1687.
- [36] O. Mayor, J. Bonada, and A. Loscos, “Performance analysis and scoring of the singing voice,” in *Proc. 35th AES Intl. Conf., London, UK*, 2009, pp. 1–7.
- [37] E. Molina, E. Gómez, and I. Barbancho, “Automatic scoring of singing voice based on melodic similarity measures,” *Master’s thesis, Universitat Pompeu Fabra, Barcelona, Spain*, pp. 9–14, 2012.
- [38] R. Sridhar and T. Geetha, “Swara indentification for south indian classical music,” in *Information Technology, 2006. ICIT’06. 9th International Conference on*. IEEE, 2006, pp. 143–144.
- [39] J. Salamon, E. Gómez, D. P. Ellis, and G. Richard, “Melody extraction from polyphonic music signals: Approaches, applications, and challenges,” *IEEE Signal Processing Magazine*, vol. 31, no. 2, pp. 118–134, 2014.
- [40] J. Serra, G. K. Koduri, M. Miron, and X. Serra, “Assessing the tuning of sung indian classical music.” in *ISMIR*. Florida, 2011, pp. 157–162.
- [41] P. Chordia, M. Godfrey, and A. Rae, “Extending content-based recommendation: The case of indian classical music.” in *ISMIR*, 2008, pp. 571–576.
- [42] G. K. Koduri, S. Gulati, P. Rao, and X. Serra, “Rāga recognition based on pitch distribution methods,” *Journal of New Music Research*, vol. 41, no. 4, pp. 337–350, 2012.
- [43] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” *IEEE Transactions on speech and audio processing*, vol. 10, no. 5, pp. 293–302, 2002.

- [44] A. Krishnaswamy, “Application of pitch tracking to south indian classical music,” in *Multimedia and Expo, 2003. ICME’03. Proceedings. 2003 International Conference on*, vol. 3. IEEE, 2003, pp. III–389.
- [45] D. Kim, K.-s. Kim, K.-H. Park, J.-H. Lee, and K. M. Lee, “A music recommendation system with a dynamic k-means clustering algorithm,” in *Machine Learning and Applications, 2007. ICMLA 2007. Sixth International Conference on*. IEEE, 2007, pp. 399–403.
- [46] “Physics of music-notes,” <http://pages.mtu.edu/~suits/NoteFreqCalcs.html>, accessed: 2018-11-30.
- [47] D. J. Berndt and J. Clifford, “Using dynamic time warping to find patterns in time series.” in *KDD workshop*, vol. 10, no. 16. Seattle, WA, 1994, pp. 359–370.