TOPICS IN

COMPUTER MUSIC

Automatic Raga Recognition in Hindustani Classical Music

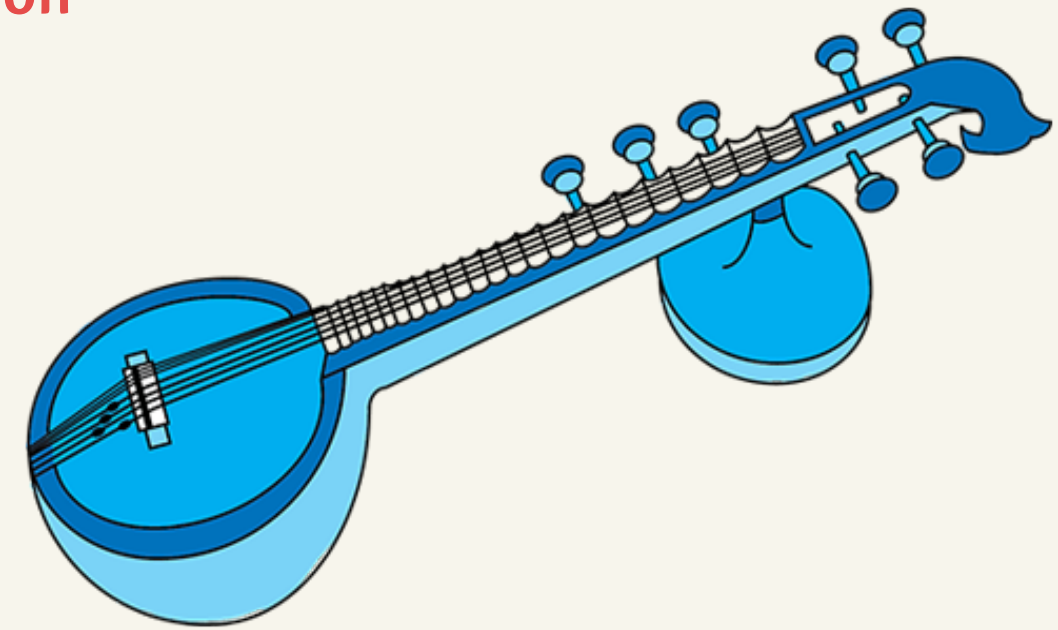
Sanchit Alekh

31st May 2017

sanchit.alekh@rwth-aachen.de

HPAC AIces RWTH AACHEN UNIVERSITY

OUTLINE

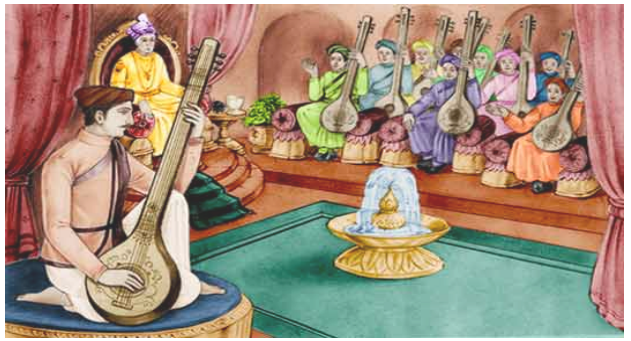


INTRODUCTION

1. What is Hindustani Music?
2. Differences with Western Music
3. Common Instruments
4. What is a Raga?
5. Why is Raga Recognition challenging?

WHAT IS HINDUSTANI MUSIC?

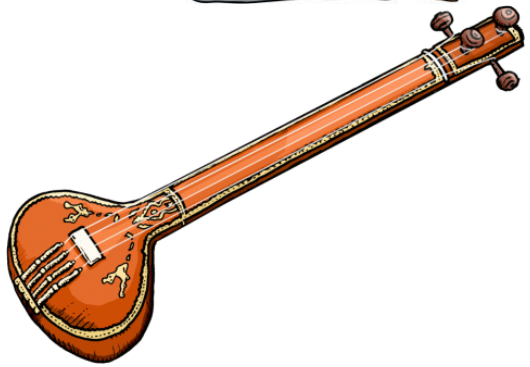
- Ancient form of music originating in Northern India
- Existence has been known and documented since 1500 B.C.
- Broadly classified into Vocal and Instrumental styles
- Traditionally passed down the generations by a ‘guru’
- Characterized by imagination and improvisation



COMMON INSTRUMENTS

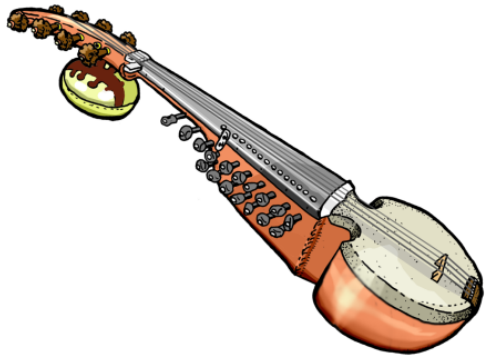
Sitar (सितार)
Stringed Instrument used mostly in solo performances

Swarmandal (स्वर्मंडल)
Stringed Instrument used as a drone / accompaniment



Tanpura (तानपुरा)
Stringed Instrument used as a drone / accompaniment

Tabla (तबला)
Percussion instrument Keeps rhythm and cyclic time signatures



Sarod (सरोद)
Stringed Instrument (fretless) used in solo performances

Shehnai (शहनाई)
Wind instrument similar to an oboe for solo renditions

WHAT IS A RAGA?

- A collection of melodic gestures, along with techniques for developing them[2]
- Compositions are bound to a rhythmic cycle*
- Longer phrases can be built by joining these melodic atoms together[2]
- Ragas are always associated with a time of the day to be sung in (sometimes the season) based on the mood they invoke

(७६)

श्यामकल्याण–त्रिताल (द्रुत).

स्थायी.

गम पध मंप गम	पग म रे सा	रे – सा सा	रे – नि़ सा
श्याऽ ऽऽ ऽऽ ऽऽ	ऽऽ ऽ म क	ल्या ऽ ऽ ण	गा ऽ व त
०	३	×	२
मं मं प प	मं प ध प	ध – – मं	प – म ग
नि स दि न	च तु र गु	नी ऽ ऽ ऽ	ऽ ऽ ऽ ऽ।
०	३	×	२

अन्तरा.

प – सां सां	सां – सां –	प नि सां रें	सां नि ध प
मे ऽ ल क	ल्या ऽ षी ऽ	अं ऽ श ख	र ज क र
०	३	×	२
मं – प प	मं प ध प	ध – – मं	प – म ग
म ऽ ध्य म	जु गु ल गु	नी ऽ ऽ ऽ	ऽ ऽ ऽ ऽ।
०	३	×	२

श्यामकल्याण–झपताल (मध्यलय)

स्थायी.

ग					रे
प	–	मग म रे	सा	रे	नि़ सा सा
सू	ऽ	नोऽ ऽ अ	हो	ऽ	श्या ऽ म
×		२	०		३
प मं	प मं	प – ध	प	ध	ध मं प मग
इ	त	नी ऽ मो	री	बि	नं ऽ तीऽ।
×		२	०		३

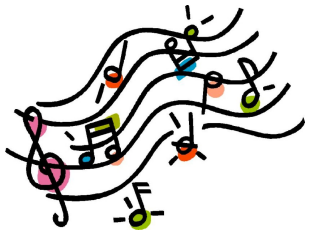
*Alap is a notable exception

ELEMENTS OF A RAGA PERFORMANCE

SOME RAGAS - #1

RAGA BHIMPALASI

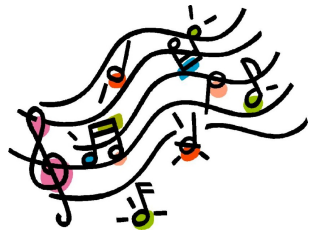
- Time : Afternoon
- Ascent : नि सा ग म प नि सां
  In C : (A# C D# F G A# C)
- Descent : सां नि ध प म ग रे सा
  In C : (C A# A G F D# D C)
- **Stressed Notes :** म (F) and सा (C)
- **Emotions :** Happiness, Pleasure



Performed on: Vocals
Performed by: Kaushiki Chakrabarty

RAGA CHARUKESHI

- Time : Late Morning – Noon (9:00 – 12:00)
- Ascent : सा रे ग म प ध नि सां
  In C : (C D E F G G# A# C)
- Descent : सां नि ध प म ग रे सा, ध नि सा
  In C : (C A# G# G F D# D C, G# A# C)
- **Stressed Notes :** सा (C) and म (F)
- **Emotions :** Pathos , Devotion , Beauty

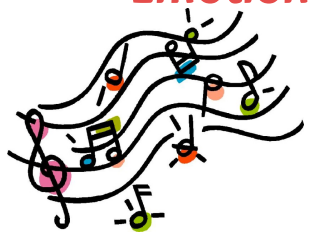


Performed on: Sitar
Performed by: Ustad Shahid Pervez

SOME RAGAS - #2

MIYA KI MALHAR (Emperor's Malhar)

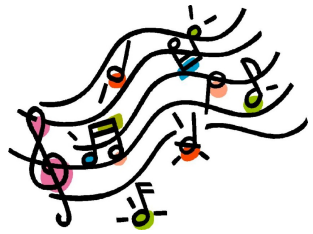
- **Time :** Any time in the monsoon, otherwise at night
- **Ascent :** सा रे म रे प, नि ध नि सां
  In C : (C D F D G, A# A B C)
- **Descent :** सां नि प, म प म रे सा
  In C : (C A# G, F G F D C)
- **Stressed Notes :** म (F)
- **Emotions :** Rains, Clouds, Thunder, Love



Performed on: Sitar
Performed by: Budhaditya Mukherjee

RAGA JOG

- **Time :** Late Night
- **Ascent :** सा ग म प नि सां
  In C : (C E F G A# C)
- **Descent :** सां नि प म ग म, सा ग सा
  In C : (C A# G F E F, C D# C)
- **Stressed Notes :** म (F)
- **Emotions :** Calmness, Love



Performed on: Flute
Performed by: Pandit Ronu Majumdar

CHALLENGES FOR RECOGNITION

1. Richness, Diversity and Complexity of melodic types, making feature extraction difficult [2]

2. Prevalence of Continuous Pitch Motions and ornamentations such as vibrato (आंदोलन) and portamento (मींड)

3. Arbitrarily tuned tonic pitch, depending on the instrument and/or the voice of the vocalist

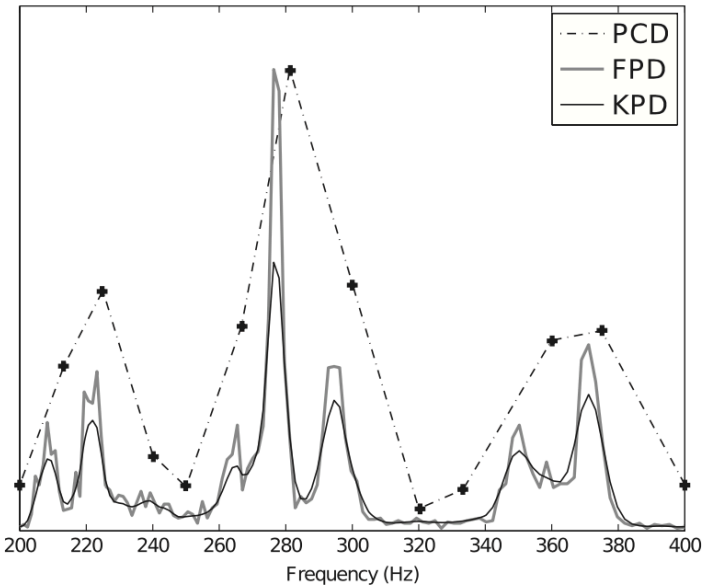
4. Subtle and intricate differences between a few ragas introduces ambiguity in recognition

APPROACHES FOR RAGA RECOGNITION

1. Based on Pitch Distributions (Chordia et al.)
2. Using Vector-Space Models (Gulati et al.)

APPROACH 1 (Chordia et al.) AT A GLANCE[2]

- Based on Pitch Distributions (PD)
- Performs simultaneous recognition of raga and tonic
- Uses a Kernel-density pitch distribution apart from the standard 12-dimensional PCD
- Uses a Nearest-Neighbour classifier with Bhattacharya distance, attaining a 4.2% tonic rate and a 10.3% raga error rate



HPAC AIcES RWTH AACHEN UNIVERSITY

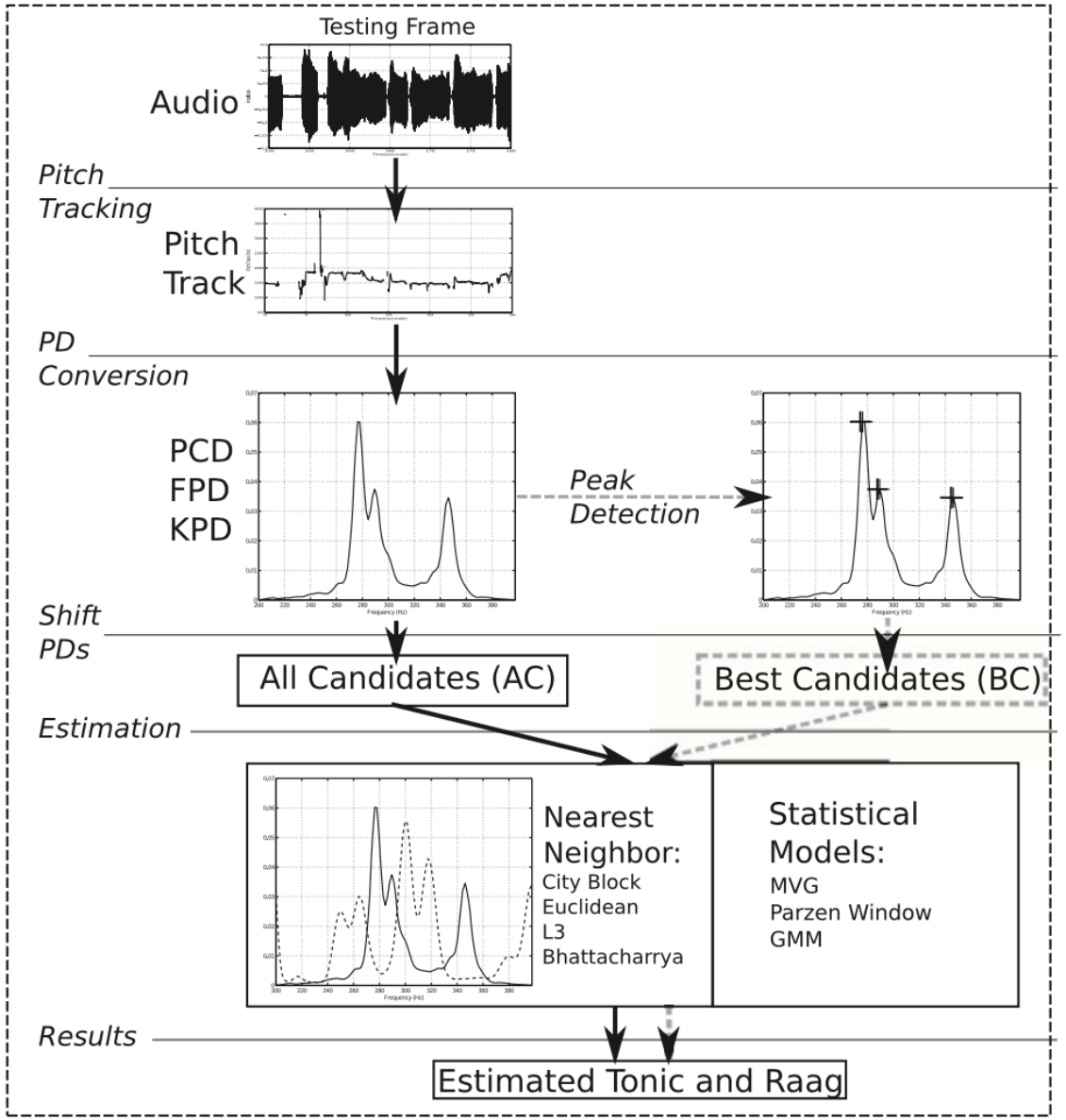
FLOWCHART

STEP - 1

STEP - 2

STEP - 3

STEP - 4



STEP 1: Pitch Tracking

1. Sawtooth-waveform-inspired pitch estimator (SWIPE')[5] is used. Pitch is estimated as the fundamental frequency of sawtooth that best matches input signal

2. Sound clip divided into 30 second chunks, after converting it to mono (to remove accompanying instruments)

3. Pitch is estimated every 10 ms, and kept within range 73.4 - 587.2 Hz using a resolution of 48 steps per octave

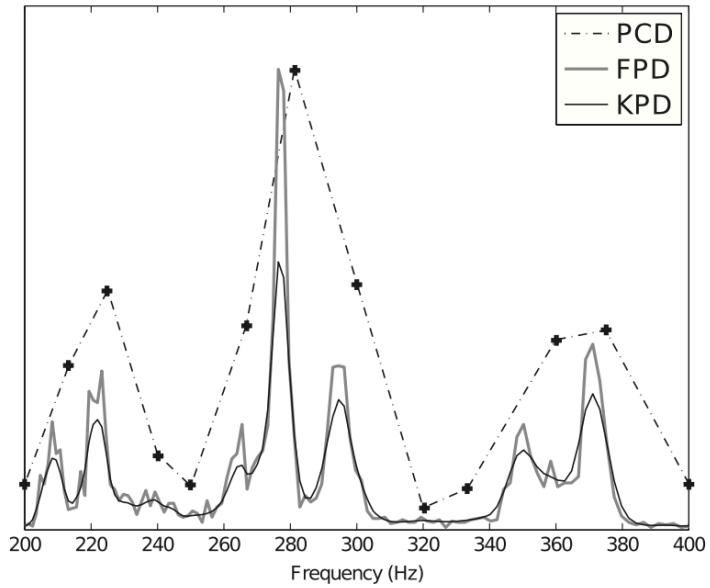
4. SWIPE' also returns an estimate of the pitch strength, which is a number between 0 and 1.

STEP 2: Extracting Tonal Features

- Three types of pitch distributions (PD) are used: a 12-D Pitch Class Distribution (PCD), Fine-Grained PD (FPD) and Kernel-Density PD (KPD)

- Each pitch estimate from the previous step is assigned in the following ways:
  - PCD assigns them to 12 chromatic pitch classes
  - FPD uses 120 or 240 bins (width of 10 or 5 cents)
  - KPD approximates a continuous PD function

- KPD centers a Gaussian window on the pitch value, and sum of all curves gives overall density



$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}}$$

where
K is the kernel with a kernel width of h,
x_i is the value of the ith pitch value, and
n is the total number of pitch values.

STEP 3: Tonic Estimation

1. Based on calculating PF for different tonic pitches and finding one that gives the best match with the database

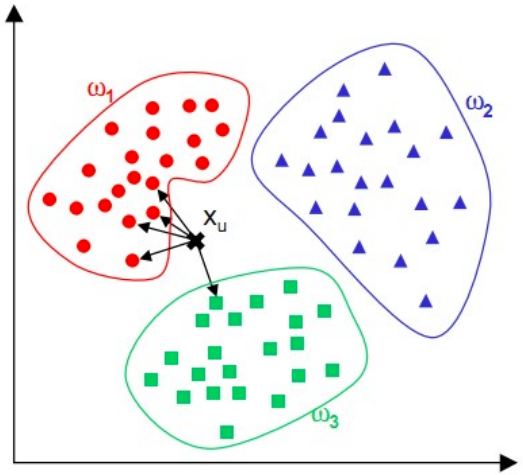
2. In All-Candidates approach, Brute Force method is used. PD is calculated for all 120/240 candidates in a circular fashion

3. The Best Candidates approach uses the fact that stable notes have peaks in HPDs. Only the 7 highest peaks are considered

4. PD is compared with all samples in the training database, and the nearest neighbour is found. The one with least overall distance is taken as the tonic

STEP 4: Raga Recognition

- Raga Recognition takes two alternative approaches: Nearest Neighbour Classification (k-NN) and Statistical Classifiers
- k-NN uses several alternative distance measures : City-Block, Euclidean, L3 norm and Bhattacharya distance
- Bhattacharya Distance is very popular for comparing probability densities, and is given by the formula
- For the statistical approach, Bayes' Rule[6] is used
- For estimating the prior probabilities, multivariate Gaussian models are used with maximum-likelihood approach



$$D_B(p,q) = -\ln\left(\sum_{i=1}^{n}\sqrt{p_i q_i}\right)$$

where $p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$.

$$P(raag_i|x) = \frac{P(x|raag_i)P(raag_i)}{\sum_j P(x|raag_j)P(raag_j)}$$

EXPERIMENTS AND RESULTS

- For our experiments, our input are audio files of classical pieces, ranging from 3-60 minutes. The database used, is called the GTraagDB*
- The expected output from the algorithm is the correct tonic and the raga
- For both tonic and raga recognition, a 10-fold cross-validation is used
- The configurable parameters for the experiments were the precision (for tonic estimation), granularity, distance algorithm, PD used, all/best candidates
- Minimum error rate of 8.5% was attained using KPD with 5-cent granularity, 15% precision Nearest-neighbour with Bhattacharya Distance & all-candidates

*paragchordia.com/data/GTraagDB

EXPERIMENTS AND RESULTS

	PCD			*FPD*			*KPD*		
	Ground	*All*	*Best*	*Ground*	*All*	*Best*	*Ground*	*All*	*Best*
City Block	29.17	46.00	39.17	21.00	26.50	44.33	22.67	25.67	36.00
Euclidean	29.33	49.17	41.67	30.00	36.50	62.33	32.33	37.00	50.00
L3	32.50	48.83	43.33	39.50	45.67	71.67	40.17	46.33	58.83
Bhattacharyya	12.50	30.00	21.83	8.67	14.50	33.50	8.50	12.50	19.17
MVG	35.17	57.67	50.83	26.00	37.00	51.83	35.17	44.67	44.17
Parzen	33.50	53.17	46.67	27.50	34.17	53.00	33.67	42.33	43.00
GMM	31.67	43.83	39.50						

"Ground" is the error rate when the tonic is known in advance; "All" and "Best" are the all-candidates and best-candidate methods, respectively.

	Ground Truth		*All Candidates*		*Best Candidates*	
	10 Cents	*5 Cents*	*10 Cents*	*5 Cents*	*10 Cents*	*5 Cents*
30-sec frames	19.09	18.13	26.45	26.26	37.62	35.70
60-sec frames	11.10	12.33	15.85	15.95	27.80	22.93
120-sec frames	8.50	10.85	12.50	11.86	19.17	18.64

Error rates calculated using KPD with NNB and 15-cent precision.

Results for Raga Recognition for different configurable parameters

CONCLUSION AND SCOPE FOR FUTURE WORK

- Evidence that melodic estimation is possible with good accuracy in a complex musical genre with continuous pitch movements and diversity of scale types
- Rich, fine-grained pitch distributions performed significantly better than PCD. Out of all methods, k-NN classification with Bhattacharya distance easily outperformed the others
- Some ragas which were misclassified, e.g. Desh and Khamaj, Asavari and Darbari etc. are sometimes difficult even for seasoned listeners
- Modeling sequential information using HMMs can be an improvement for this approach. N-gram modeling is also a promising option

REFERENCES

- [1] S. Chakraborty, G. Mazzola, S. Tewari, and M. Patra, 'Computational Musicology in Hindustani Music'.
- [2] P. Chordia and S. Sentürk, 'Joint Recognition of Raag and Tonic in North Indian Music'
- [3] S. Gulati, J. Serra, V. Ishwar, S. Sentürk, and X. Serra, 'Phrase-based raga recognition using vector space modeling'
- [4] Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 'Fast unfolding of communities in large networks'
- [5] Camacho Lozano, Arturo. 'SWIPE: A sawtooth waveform inspired pitch estimator for speech and music.'
- [6] Duda, Richard O., Peter E. Hart, and David G. Stork. 'Pattern classification'. John Wiley & Sons, 2012.

THANK YOU FOR YOUR ATTENTION

QUESTIONS ?

WHAT IS A RAGA?







Video Link for PDF Version: https://goo.gl/photos/jWR17cxJnUDpdSd77

DIFFERENCES WITH WESTERN MUSIC[1]

Western Classical Music

1. Concerts are never performed as an extempore
2. Percussion is not very important
3. The tonic never changes. E.g. a composition in C minor will always be performed in C minor
4. Major-Minor tonal systems, harmonies and counterpoints are important

Hindustani Music

1. The composition is always improvised upon. There is no script
2. Percussion is extremely important
3. The pitch of the tonic changes according to the instrument/voice but relative distances are constant
4. Harmony is not emphasized, there are no counterpoints, generally a single-melody instrument/voice

APPROACH 2 in BRIEF

- Uses the vector-space modeling analogy from Text Information Retrieval
- Motivation behind the approach is the way seasoned listeners identify ragas
- Melodic patterns from a collection of audio recordings in an unsupervised way
- Similar patterns are grouped, a directed graph is created based on similarity thresholds and clustered using a Community Detection Method[4]
- Each recording is represented as a Vector, frequency and inverse frequencies are extracted and fed to a classifier such as SVM/Logistic Regression