Computational Musicology for Indian Music

Preeti Rao

Department of Electrical Engg., IIT Bombay

WiSSAP 2023
I.I.T. Kanpur

Universals in Music

Melody: Music tends to use discrete pitches to form scales, containing 7 or fewer scale degrees per octave.

Rhythm: Music tends to use an isochronous beat organized according to metrical hierarchies based on multiples of two or three beats. This beat can be used to construct motivic patterns.

Form: Music tends to consist of short phrases, less than 9 s long.

Instrumentation: Music tends to use both the voice and (nonvocal) instruments, often together in the form of accompanied vocal song.

Performance style: Many different performance styles.

Outline

- Music Concepts: Melody and Rhythm
- Describing a Concert
- Identifying Interesting Problems (and Data)

Making Music

Classifying instruments (*Natyashastra*) by the mode of sound production

- Stretched strings
- Membranes
- Solids
- Hollow (air)

Above all, is the “Human Body” instrument

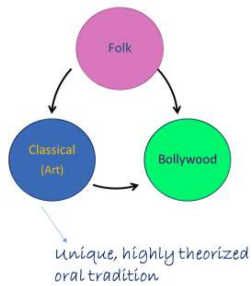
- the most expressive instrument

Supremacy of the singing voice is characteristic of many cultures.

Vocal Music Diversity

4

The Music of India



Classical Music Map of India: ESRI

Hindustani and Carnatic traditions are both based on the theory of Raga (melody) and Tala (rhythm).

Summarising a piece of music in terms of musical aspects such as the melody, harmony, rhythm, texture, dynamics….

The music itself could be the written score or an audio signal.

Multiple applications:

- Musicology (empirical analyses of music archives)
- Music recommendation (categorizing music by recognizing similarity)
- Music creation (synchronization).

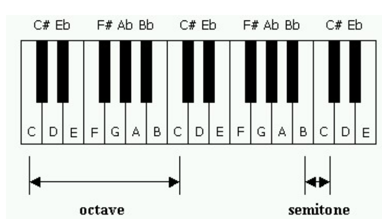
Music theory: Scales

The list of all the notes that are expected or allowed in a particular piece of music is a scale. E.g. major scales. All major scales (keys) sound the same. Music can be transposed from one key to another in major scales.

Some traditions use 'modes'. The mode, like a scale, lists the notes that are used in a piece of music. The many modes are different from each other in terms of the intervals between the 'scale' notes. Music cannot be transposed across modes.

Equal temperament

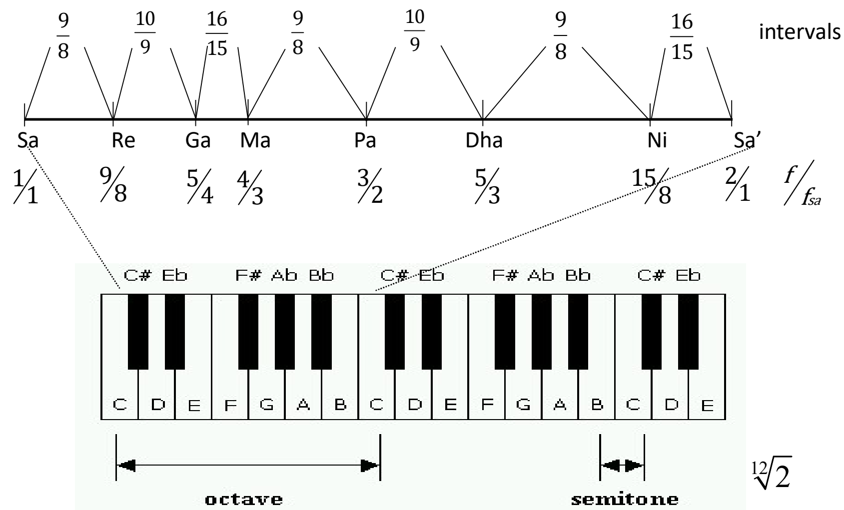
$$x(cents) = 1200\ x\ log2\ \frac{f_2}{f_1}$$



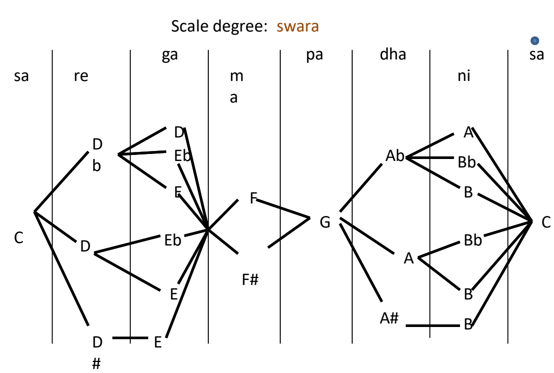
Ratio = $\sqrt[12]{2}$

8

A common musical scale: the "just intonation" scale



Parent scales or modes (Melakarta)*



Titon, Worlds of Music

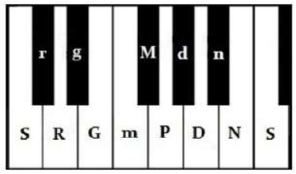


Figure 1: The solfege of Hindustani music shown with an arbitrarily chosen tonic (S) location.

		Re(2)		Ga(3)		ma(4)				Dha(6)		Ni(7)
Sa(1)	+	or	+	or	+	or	+	Pa(5)	+	or	+	or
		re(2♭)		ga(3♭)		Ma(4#)				dha(6♭)		ni(7♭)

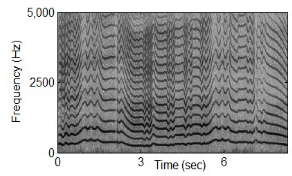
The North Indian Classical Music Performance

Concert sections
- alap
- bandish (vistaar)
- taan

all within the chosen melodic (**raga**) and rhythmic (**tala**) framework

Signal characteristics

Vocalist (main melody)



Tanpura (drone)

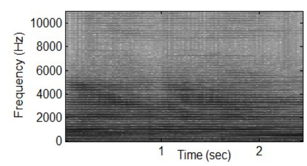
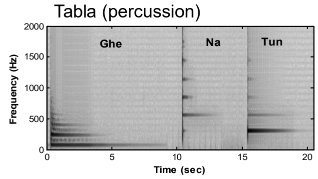
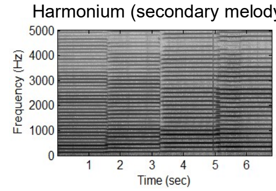


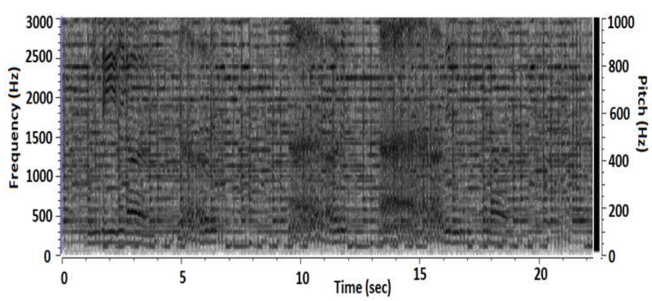
Tabla (percussion)



Harmonium (secondary melody)



The music

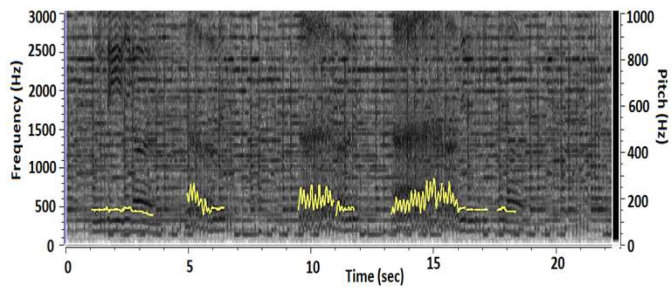


Bhimsen Joshi, Marwa, Tintal
Bandish: Guru Bina Gyan Na
Pave

The “melody”

‘Predominant pitch’ detection at 10 ms intervals

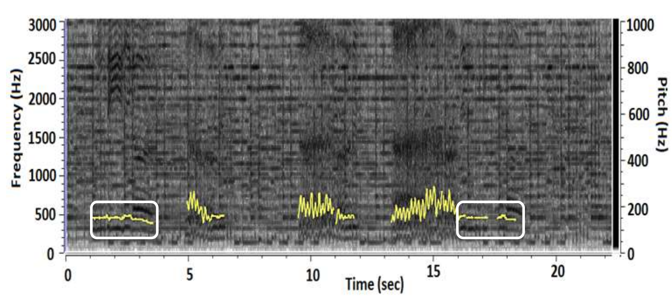
F0 contour



Bhimsen Joshi, Marwa, Tintal
Bandish: Guru Bina Gyan Na Pave

The “melody”

melodic motif



Bhimsen Joshi, Marwa, Tintal
Bandish: Guru Bina Gyan Na Pave

17

Hindustani (North Indian) Music Performance

- Many instruments
- Many artists and playing styles
- Many ragas or melodies
- **Similar concert structure**

- The performance is a process with affective powers through the raga (melody) that is revealed. Indian music rhythm is organized at the level of cyclic rhythmic patterns (tal or metre).
- The stages of a concert progress through a gradual exposition, development, and acceleration (stemming probably from the historical role of music as a religious ritual).
- Tempo and rhythmic density define the large-scale organization of the concert as well as the episodes of improvisation.

18

Dhrupad and Instrumental Hindustani Music

Dhrupad vocal

- Lead - Vocals
- Accompaniment - *Pakhawaj*
- *Tanpura*, supporting vocals

(https://www.youtube.com/watch?v=Icgcgya3qM4)

Sitar

- Lead - Sitar
- Accompaniment - *Tabla*
- *Tanpura*

(https://www.youtube.com/watch?v=FnP1x5FuEqU)

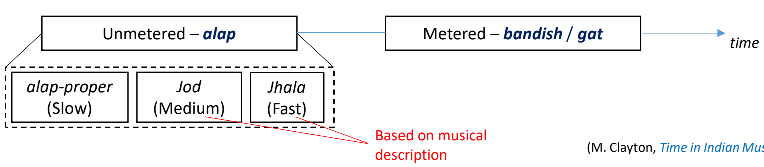
Concert Structure

- Typical concert
  - 1.5 hours long
  - Consists of *raga* performances

Section durations missing

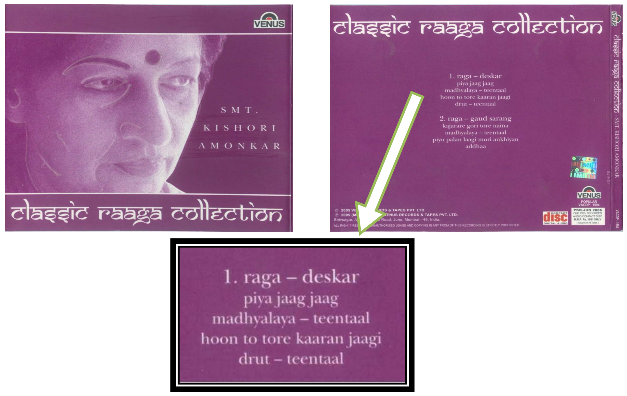
(http://www.nirmalyadhrupad.org/media)

- Typical *raga* performance



(M. Clayton, *Time in Indian Music*, 2000)

Common available cover information



Music Information Retrieval

Mostly work in audio signal processing addressing research in:

- Automatic music transcription
- Music recommendation
- Style characterization and music tagging
- Music summarization and thumbnails
- Empirical analyses of music data for musicology studies

The problem statements and computational tools are culture dependent.

Identifying relevant tasks

	Approach	Example tasks	Uses
Top-down	Search audio recordings for music-theoretical concepts.	*Raga* and *tala* identification	Help organise large corpora and save time-consuming work of identifying items
Bottom-up	Ask what patterns and invariants could be discovered in the music in a data-driven fashion	Pattern discovery, in melody or rhythm, from first principles	Reveal important patterns not otherwise recognised by theory, potentially revealing (for example) cognitive processes.
Critical	Explore the gaps between theoretical concepts and practice.	How adequately do raga definitions describe musical practice?	Improve music theory by testing it against practice

From: M. Clayton, Hindustani rhythm and computational analysis: A musicological perspective, Indian Art Music: A Computational Perspective, Eds: P. Rao, H.A. Murthy, SRM Prasanna, 2022.

23

Motivation: Music theory vs Empirical studies

Theory is an important part of the Indian art musical traditions (many texts, treatises).

R.N. Iyengar (Invited talk in ICPR Seminar on Science & Technology in the Indic Tradition, 2017) reviewed the Sanskrit texts that discuss music theory to find that there is a large technical vocabulary used but the terms are “vague and not quantifiable”.

Empirical studies of music performance by practitioners (who have internalized the grammar and aesthetics) can help to close the gap.

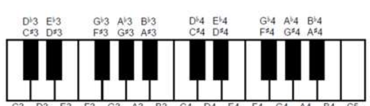
Computational tools can enable such musicological studies to scale and contribute to improving music theory by testing it against practice. Also assist in studies on style discrimination and the historical development of performance practice.

Example: Melodic features

- How to tell the key of a piece? (How do listeners identify the key?)
- An influential view of key-finding is the distributional view: the perception of key depends on the distribution of pitch-classes in the piece.

Temperley, D. and Marvin, E.W., 2008. Pitch-class distribution and the identification of key. Music Perception, 25(3), pp.193-212.

Key profile as a cognitive template



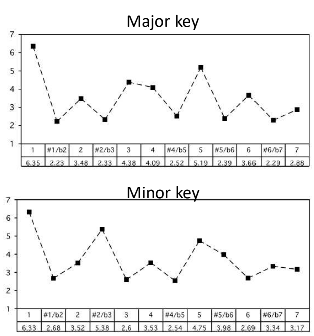


FIGURE 2. Key-profiles for major keys (above) and minor keys (below). From Krumhansl and Kessler (1982).

The tonic pitch is rated most highly, followed by other notes of the tonic triad

Krumhansl and Kessler, 1982

K-S key finding algorithm

Pitch-class distribution (PCD) = 12 D vector representing the total duration of each pitch-class in the piece.

Can be computed from the music score. Or, for audio, via transcription.

The Krumhansl-Schmuckler algorithm: the distribution of pitch classes in a piece is compared with the ideal distribution or "key profile” for each key.

Score example

PCD

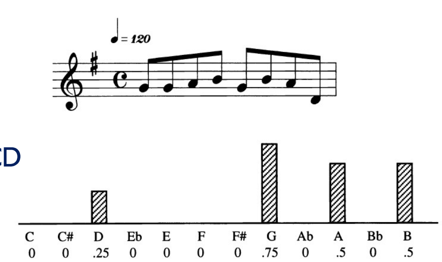


Fig. 2. Measure 1 of “Yankee Doodle,” with input vector showing total duration of each pitch class.

Temperly, 1999

Raga description

A raga is described by

- Musical characteristics
  - The choice of notes, their order and hierarchy
  - The intonation and ornamentation of specific notes
  - The recurring characteristic phrases (pakad)

- Extra-musical characteristics
  - Affect
  - Time of day
  - …

29

Tala

- The main percussion instruments used in Hindustani classical music are the tabla and (the somewhat less common) pakhavaj. The tabla is a set of two drums of different sizes.

- A tala is a cyclical rhythm pattern. It has a specified number of beats and a named tabla stroke for each beat.
- Example: 16-beat Teentala: dhaa dhin dhin dhaa / dhaa dhin dhin dhaa / dhaa tin tin taa/ taa dhin dhin dhaa

- Singing in Ektala

https://youtu.be/TgxkGOvIAOo

Music theory: raga grammar

The 12 tones within an octave:
S| r| R| g |G |m |M| P |d |D |n |N |S

"Distributional"

Bhupali
Tonal material: SRGPD
Ar: SRG, PDS *Av*: SDP, GDP, GRS
Vadi: G, *Samvadi*: D
Phrases: RDS, RPG, PDS, SDP, GDP, GRS
Natural shrutis of R, G, D

tonal hierarchy

31

Music theory: raga grammar

The 12 tones within an octave:
S| r| R| g |G |m |M| P |d |D |n |N |S

"Structural"

Bhupali
Tonal material: SRGPD
Ar: SRG, PDS *Av*: SDP, GDP, GRS
Vadi: G, *Samvadi*: D
Phrases: RDS, RPG, PDS, SDP, GDP, GRS
Natural shrutis of R, G, D

32

"Allied ragas"

Deshkar	Bhupali
Tonal material: SRGPD	Tonal material: SRGPD
Ar: SGPD, SPDS *Av*: S, PDGP, DPG(R)S	*Ar*: SRG, PDS *Av*: SDP, GDP, GRS
Vadi: D, *Samvadi*: G	*Vadi*: G, *Samvadi*: D
Phrases: SG, G(P)DPD, P(D)SP, DGP, DPG(R)S	Phrases: RDS, RPG, PDS, SDP, GDP, GRS
Higher shrutis of R, G, D	Natural shrutis of R, G, D

Students are taught the ragas together and **warned** against confusing them.

33

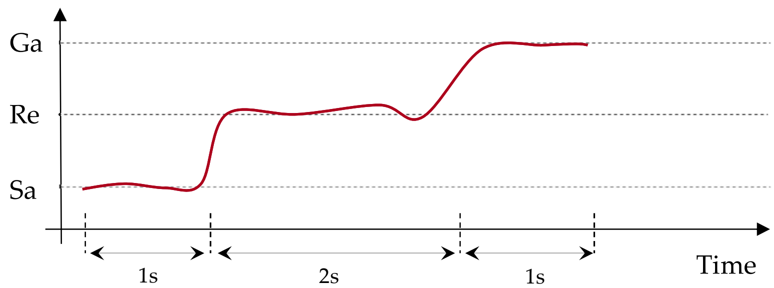
Raga identification with a PCD?

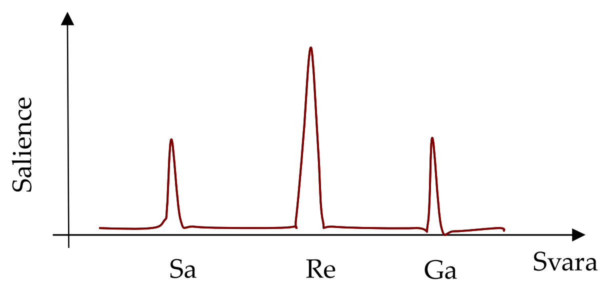
Tonal material: S R G P D

Computing PCD from audio …

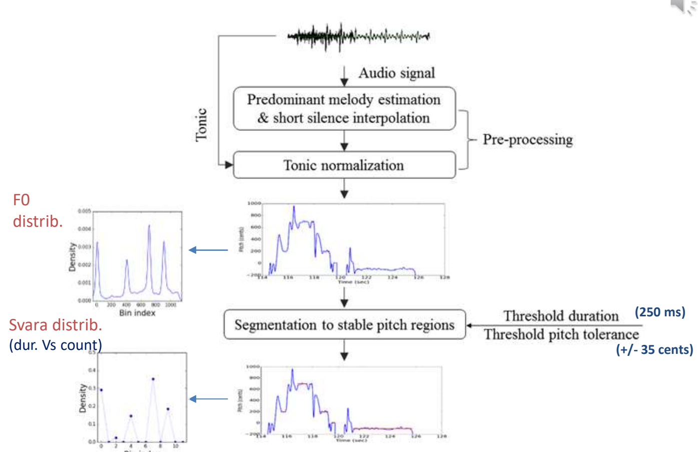
34

Pitch distribution -> raga notes





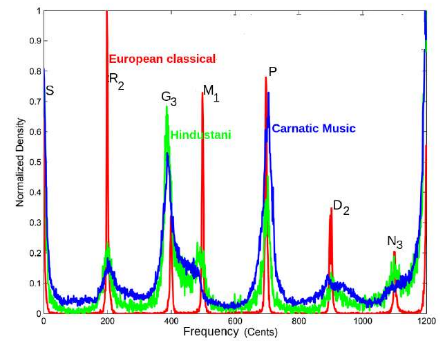
Computing the distribution



36

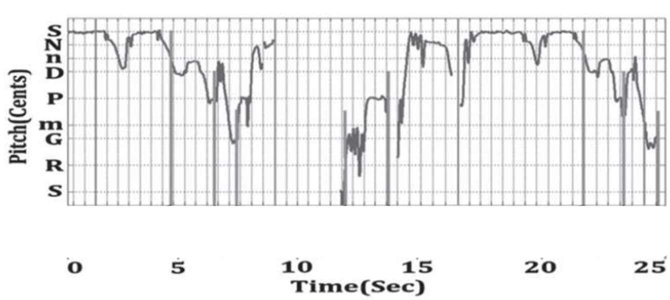
Raga recognition (like template based key detection)?

Computing the PCD from audio

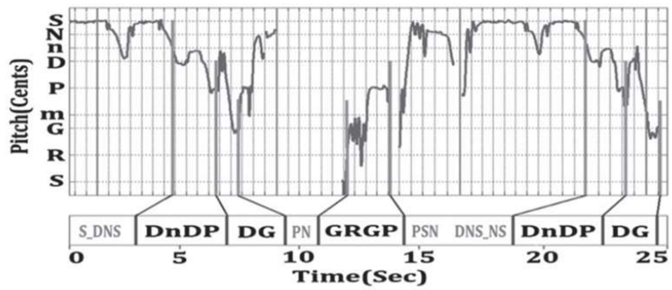


C major/Bilawal/Shankarabarana
Figure courtesy- Shrey Dutta, "Analysis of motifs in Carnatic music– a computational perspective. Master's Thesis, IITM, 2016.

F0-contour: Label the notes?



Need raga knowledge!

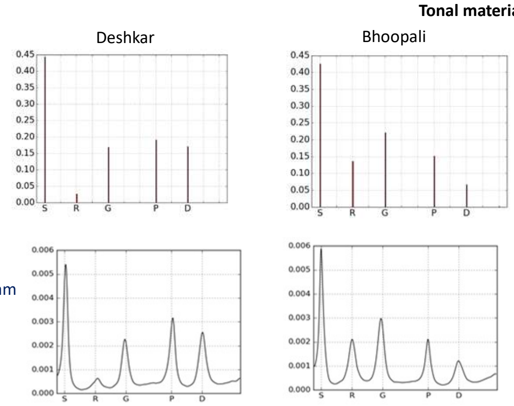


Raga identification with a PCD?

Tonal material: S R G P D

Svara salience histogram

Pitch histogram



40

*From: Ganguli, K.K. and Rao, P., Proc. ISMIR 2017

Distributional model

- **Model hyperparameters**
  - Continuous or Discrete? (I.e. Pitch or Svara = pitch class)
    - Continuous -> Pitch distribution: Bin-width
    - Discrete -> Pitch-class distribution: Svara (note) segmentation + Duration vs Count.
- **Similarity metric**
  - Correlation
  - Euclidean
  - Cityblock
  - Bhattacharya

Music theory is not precise enough.

- **Use data (labeled examples)**

Experiments: Retrieval measures

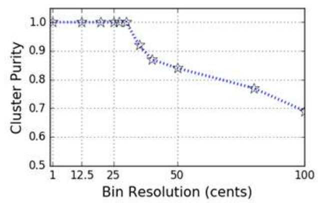
Raga	# Concerts	Duration (hours)	# Artists
Deshkar	6	2:16:50	5
Bhupali	11	5:12:23	9

- Influence of bin width (1 cent –> 100 cents) on unsupervised clustering of concert distributions. Evaluate **cluster purity** with respect to the raga discrimination.
- The chosen similarity metric is computed between distributions obtained from each of a pair of concerts to **obtain an ROC** (TPR vs FPR) for the detection of the mismatched pairs.

Cluster purity: discriminating allied ragas

CP = 1 => perfect clustering, whereas 0.5 => random clustering

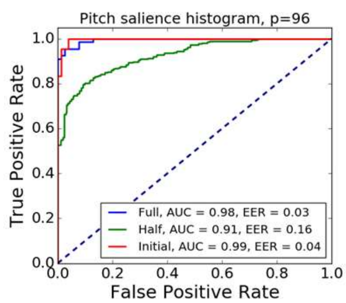
- Bin resolution for the pitch salience histograms



Pitch salience distribution → 1.0 for bin-width up to 27 cents
Svara salience histogram → 0.91, Svara count histogram → 0.84

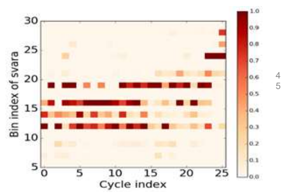
Insights: Raga discrimination performance dependence on segment

Initial: red (Initial = alap + vistaar)
Full concert: blue
Latter half: green

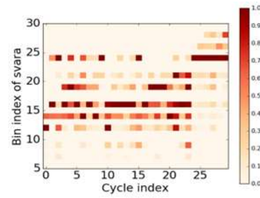


Insights: on Improvisation

Raga Bhupali : Two different performances by vocalist Pt. Ajoy Chakravarthy

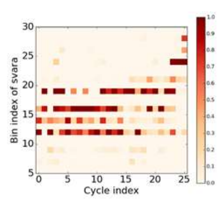


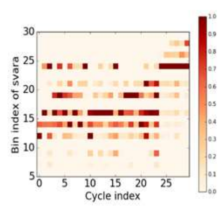
Time scale of rhythmic cycle



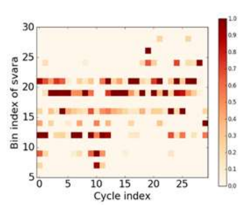
Insights: Time scale of a rhythmic cycle, same artist at diff times

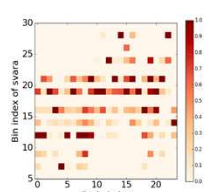
Bhupali





Deshkar





46

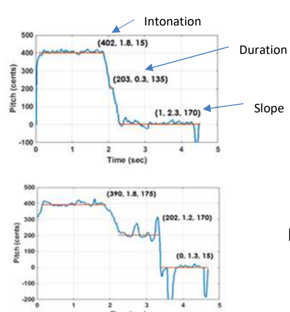
Music theory predictions?

Deshkar	Bhupali
Tonal material: SRGPD	Tonal material: SRGPD
Ar: SGPD, SPDS *Av*: S, PDGP, DPG(R)S	*Ar*: SRG, PDS *Av*: SDP, GDP, GRS
Vadi: D, *Samvadi*: G	*Vadi*: G, *Samvadi*: D
Phrases: SG, G(P)DPD, P(D)SP, DGP, DPG(R)S	Phrases: RDS, RPG, PDS, SDP, GDP, GRS
Higher shrutis of R, G, D	Natural shrutis of R, G, D

How are raga phrases (motifs) discriminated?

"GRS
"

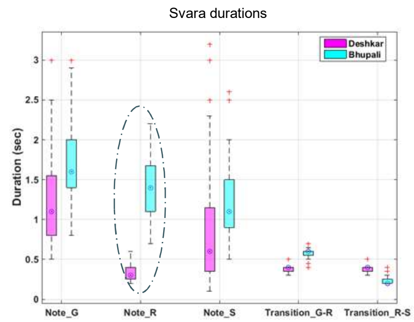
Deshkar

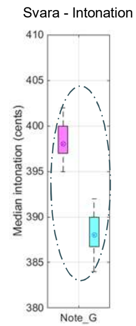


Bhupali

Phrase-based discrimination of the ragas (GRS)

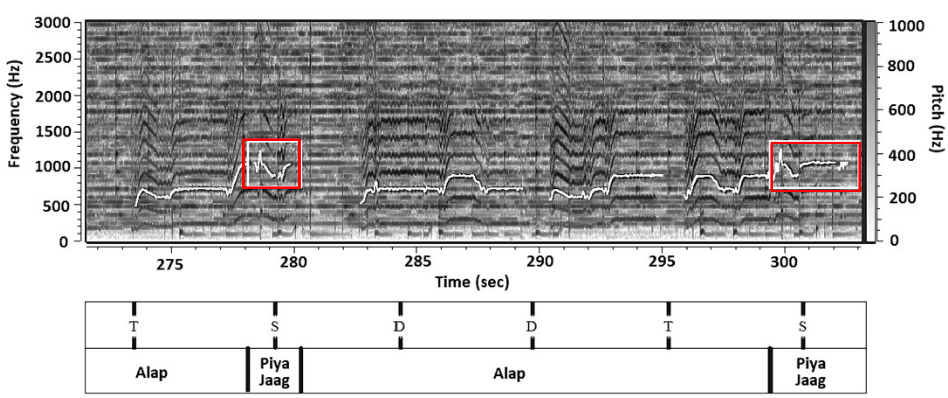
Feature selection -> R-duration, G-intonation



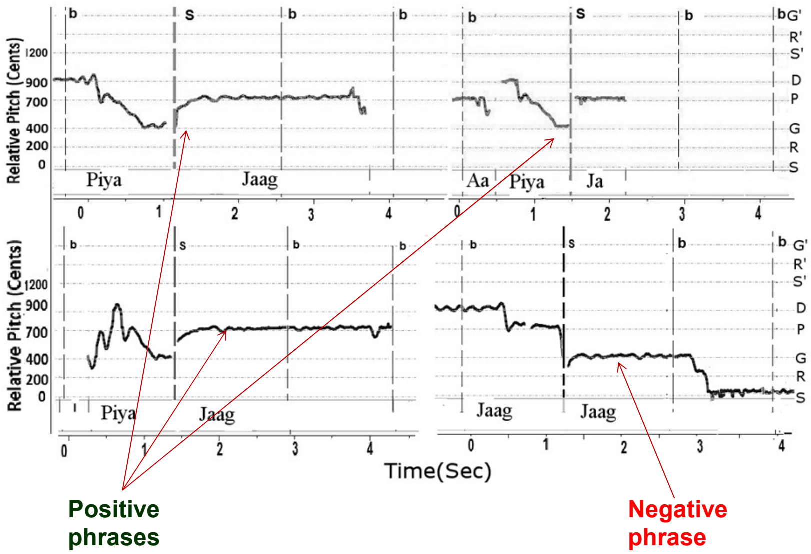


K. K. Ganguli and P. Rao "A Study of Variability in Raga Motifs in Performance Contexts", JNMR, 2021

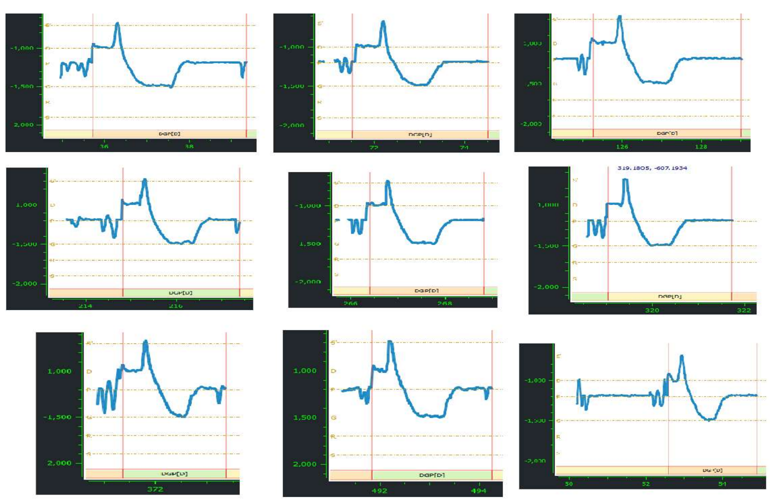
Melodic pitch contour "motif" detection



Example : 'Piya Jag' Phrases Swaras: *D P G P*



Within-concert variability of motif (DGP)



Distance measure: DTW-based pitch difference

Automatic Notation Generation in Hindustani Music

DDP Stage 1 Presentation **|** Madhumitha S **|** Guide: Prof. Preeti Rao

53

Notating Raga Music: Motivation

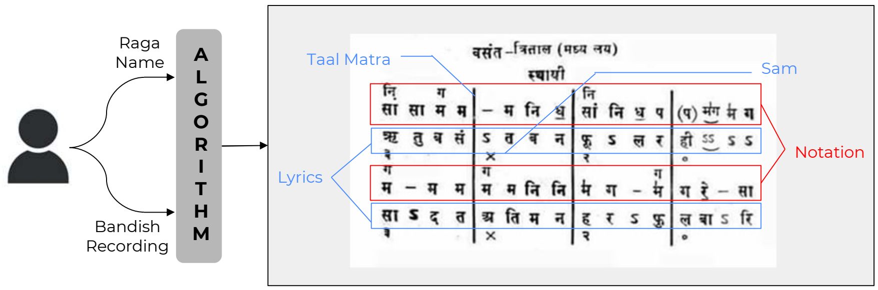
- Teaching of Hindustani Music has always been an oral tradition
- Concept of “raga”- melodic framework for improvisation in HCM
- Schematic notation is used in teaching and practice of complex compositions

An instance of sargam notation being used for music teaching

54

Problem Statement

To automatically generate schematic Bhatkhande Notation in sargam for any audio of a bandish



An excerpt of Bhatkhande notation of Bandish 'Ritu Basant' in Raag Basant

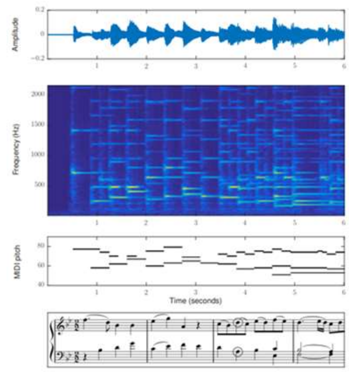
Evolution of Notation in Hindustani Music

- Historically an oral tradition
- 20th century: Pt. Bhatkhande and Pt. Paluskar notation systems using *sargam*
- Skeletal framework, strips off performance-subjective details. Became accepted and widely used in music pedagogy and preservation
- Pt. Bhatkhande compiled manual annotations of about 1,200 compositions in Marathi in six parts of the Kramik Pustak Malika series
- 10k+ bandishes, no digitally usable notations for most newer compositions
- Applications in music accompaniment generation

56

Automatic Music Transcription (AMT)

- Design of algorithms to convert acoustic music into some form of symbolic representation
- Musical equivalent of Automatic Speech Recognition (ASR)
- Involves pitch estimation, onset and offset detection, beat and rhythm tracking, score typesetting
- Extensive work has been done on AMT in Western Music
- Methods such as NNMF (non-negative matrix factorisation) and LSTMs are common

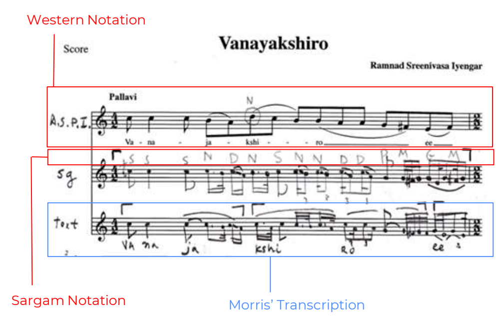


Top-to-bottom: input waveform, frequency representation, music score representation

57

Transcription v/s Notation

Morris (2011) brings out the skeletal nature of notation by transcribing a Carnatic music composition to show the different variations of each note in the notation



Variations of each note that appeared in the rendition

58

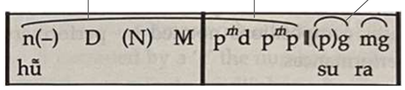
Transcription in the context of Indian Music

- Every detail observed in the pitch contour is transcribed including transients & glides
- Performance-subjective; helpful in music appreciation but lacks reproducibility
- Magriel in **The Songs of Khayal** (Vol 1 & Vol 2) transcribes several khayals with complex nomenclature (an extension of sargam notation) as shown in the below figure

[N D N M P] is a quintessential statement of Raga Hamir. Here it is expressed using two varieties of *mind*: the first slide from Ni to Dha begins with a stable Ni; the second simply touches Ni after a rapid transition from Dha before sliding down to tivra Ma, the *mind* occupying a full matra.

The phrase ends with a *khaṭka* which is commonly described by musicians as [pdpp] performed evenly in the space of two mātrās

Ga is slid to from Pa after around half a matra. Ma and Ga each occupy half a matra. The transition between them is too rapid to warrant notating as a slide



Composition in Raag Hamir notated by Nicolas Magriel

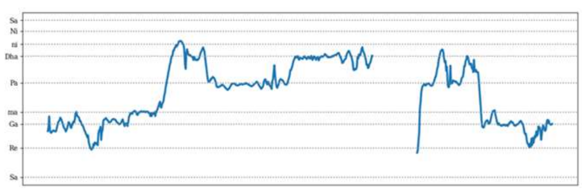
59

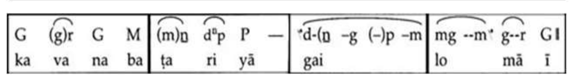
Signal Transcription v/s Notation

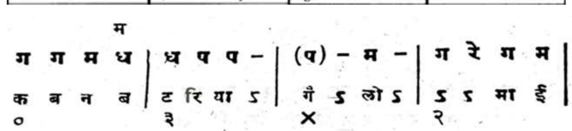
Pitch Contour

Transcription

Notation

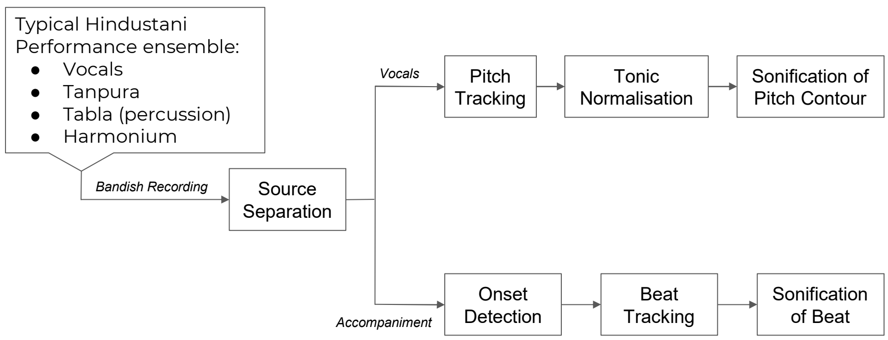






60

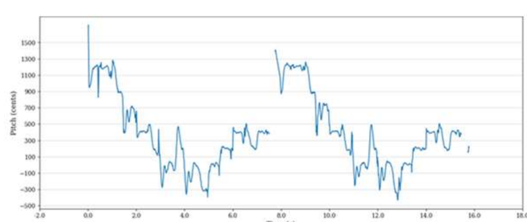
Processing Pipeline



Pipeline showing methods used to transform bandish to pitch contour with beat positions

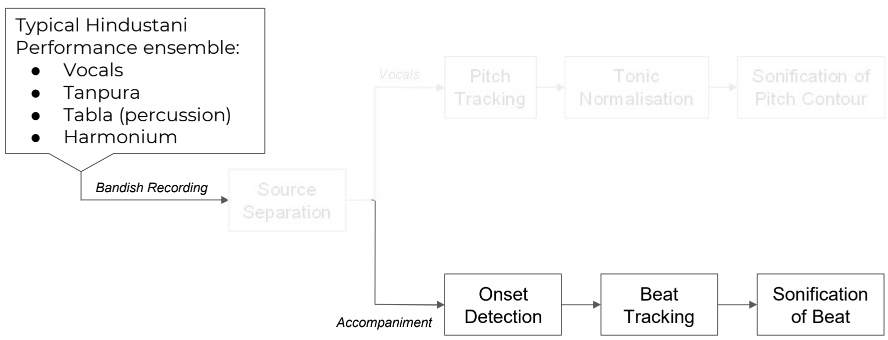
Pitch Tracking

- Pitch tracking or pitch contour is the F0 contour of the dominant vocals
- Parselmouth (Praat) pitch tracker based on autocorrelation function
- Normalisation of pitch contour w.r.t. tonic of performer
- Sonification using a sinusoid to validate pitch tracking accuracy



62

Processing Pipeline



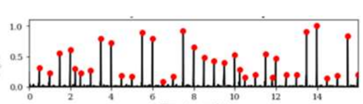
Pipeline showing methods used to transform bandish to pitch contour with beat positions

Rhythmic Features

1. Rhythm/ *taal* plays an important role in Hindustani Music
2. Notation contains demarcations of *sam*, *khaali* and *matras* of the taal
3. Each matra of the taal then contains one or more swaras

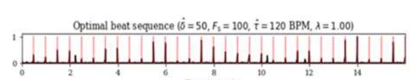
Onset Detection

- Spectral-based novelty function to find location of onsets in accompaniment
- Peak picking algorithm to select local maxima in novelty function



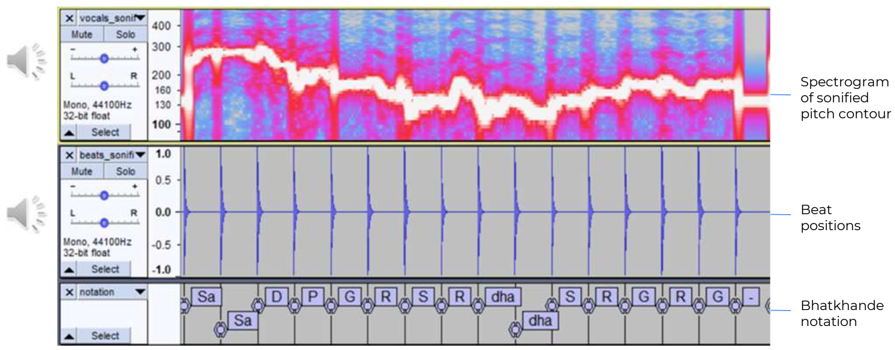
Beat Tracking

- Novelty function onsets and the assumption of a fairly constant tempo are used to determine beat positions
- Dynamic programming approach
- Sonified beats obtained through generation of clicks at beat positions



64

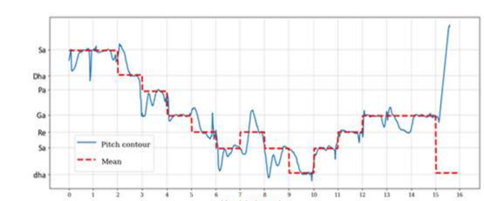
Manual Annotation on Audacity



Audacity layers used for manual annotation of bandish recordings

Results: Mean-value based Notation

For time *T*, pitch values *P* and beat positions *B*, the following algorithm was devised-



True note	Ṡ	Ṡ	D	P	G	R	S	R	Ḍ	Ḍ	S	R	G	R	G	-
Generated note	Ṡ	Ṡ	D	P	G	R	S	R	S	Ḍ	S	R	G	G	G	-

Accuracy = $\frac{N_{correct}}{N_{total}} \times 100$ = 87.5%

Algorithm:

1. For each beat interval from t_i and t_{i+1}, calculate mean pitch M_i as follows

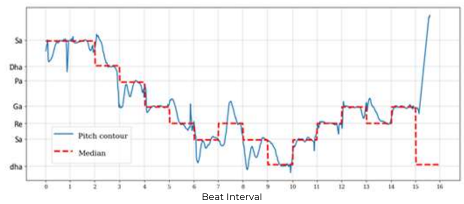
$$M_i = mean\left(\left\{\{P_j \mid T_j \in [t_i, t_{i+1})\}\right\}\right)$$

1. Find the closest raga note to the calculated mean M_i

$$Closest\ Note_i = argmin_{r \in raga\ notes}|r - M_i|$$

Results: Median-value based Notation

For time *T*, pitch values *P* and beat positions *B*, the median algorithm is as follows-



True note	Ṡ	Ṡ	D	P	G	R	S	R	Ḍ	Ḍ	S	R	G	R	G	-
Generated note	Ṡ	Ṡ	D	P	G	R	S	R	S	Ḍ	S	R	G	R	G	-

Accuracy = $\frac{N_{correct}}{N_{total}} \times 100$ = 93.75%

Algorithm:

1. For each beat interval from t_i and t_{i+1} , calculate median pitch M_i as follows

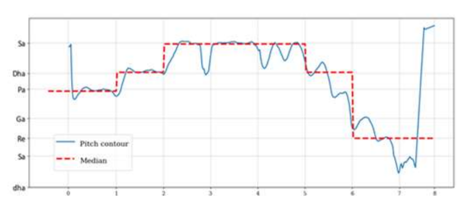
   $$M_i = median\left(\left\{\{P_j \mid T_j \in [t_i, t_{i+1})\}\right\}\right)$$

1. Find the closest raga note to the calculated median M_i

   $$Closest\ Note_i = argmin_{r \in raga\ notes}|r - M_i|$$

Challenges & Limitations

- Here, each matra has multiple notes- highlighting that a matra-level analysis may not suffice in accurately obtaining the notation
- Accuracy of the model drops to 62.5% as we have not accounted for the possibility of multiple notes in a beat
- Typically, Bhatkhande notation contains at-most 4 swaras in a matra



True note	P	D	Ṡ	Ṡ	ṠṠ	DP	GR	S
Generated note	P	D	Ṡ	Ṡ	Ṡ	D	R	S

Exploring the correspondence between singers' gestures and melody with deep learning

Sound-gesture relationship in music

- There are direct associations such as with gestures involved in playing an instrument.
- Other prominent links are the gestures in tapping and coordinating, visual imagery associated with the melody or lyrics, and representing the flow and organization of the phrases.
- Vocalists in Indian classical music traditions are known to use a wide range of manual gestures to accompany their singing.
- We study the possible complementarity between gesture and melodic movement in the context of Hindustani vocal performances.

Previous work on gesture in vocal music

- Musicological studies based on interviews with musicians and the manual analysis of videos [Rahaim, 2009].
- Synchronization and coordination in Hindustani music ensembles, i.e. gestures marking time [Clayton, 2013]
- Sound-gesture relationship in the context of Carnatic music pedagogy [Pearson, 2013]

Our goal is to develop computational methods that can be scaled to enable corpus-wide studies.

Our dataset

- North Indian Raga dataset (from IEMP Corpus of Durham University, available on OSF) of audiovisual recordings of concerts
- Multitrack audio recordings accompanied by videos from up to 3 cameras
- Valuable for empirical and computational musical analyses
- We analyse the solo music section of *alap* singing for the high-level semantic task of raga classification of short snippets.

AG_Marwa: 1:40

Dataset

We use the OSF dataset comprising of alap (2 takes) and characteristic phrases (pakad) in 9 ragas, sung by 3 professional artists (~3.5 hours total)

Raga	Scale
Bageshree (Bag)	S R g m P D n
Bahar	S R g m P D n N
Bilaskhani Todi (Bilas)	S r g m P d n
Jaunpuri (Jaun)	S R g m P d n
Kedar	S R G m M P D N
Marwa	S r G M D N
Miyan ki Malhar (MM)	S R g m P D n N
Nand	S R G m M P D N
Shree	S r G M P d N

Video stills of singers from left to right: Apoorva Gokhale (AG), Chiranjeeb Chakraborty (CC) and Sudokshina Chatterjee (SCh)

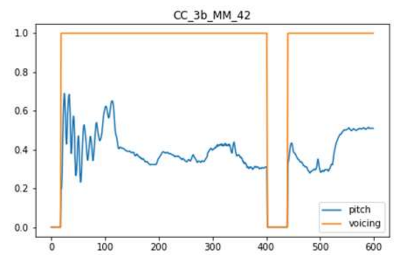
Gesture representation by pose estimation

- Use OpenPose 2D pose estimation [1] to track a set of upper-body keypoints.
- Normalize keypoints based on square bounding box around singer
- We use only left and right wrist keypoints to get time series
- Any missing data is interpolated; lowpass filter smoothing.

[1] OpenPose: Main Page (cmu-perceptual-computing-lab.github.io)

Audio Features

- Pitch data - tonic normalised F0 estimate (in cents)
- Voicing data (binary valued)

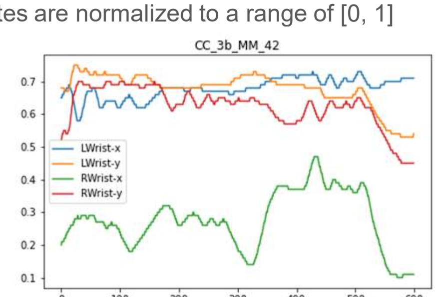


Plot of audio features for one 12s clip for 20 ms frames

Video features

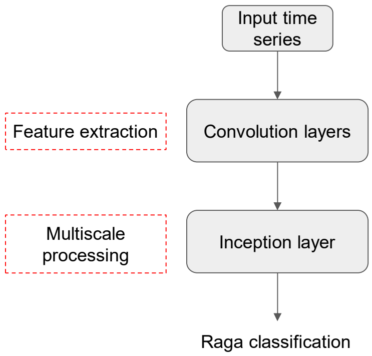
We extract the positions (x and y coordinates) of the right and left wrists from each frame of the video using OpenPose, available at 25 fps.

Note: The coordinates are normalized to a range of [0, 1]



Plot of video features for one 12s clip

Deep learning model



C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

Fawaz, Lucas, Forestier *et al.* InceptionTime: Finding AlexNet for time series classification. *Data Min Knowl Disc* **34**, 1936–1962 (2020).

Experiment 1

Unimodal audio/video raga prediction

Experiment (unimodal raga prediction)

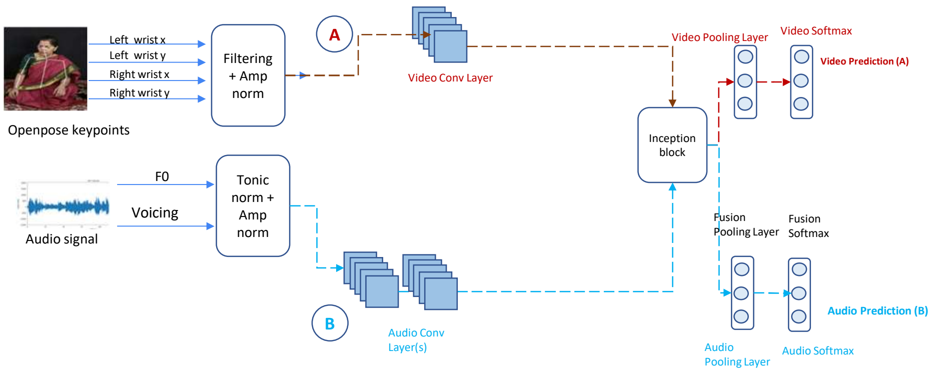


Diagram of unimodal raga prediction model with Audio (represented as A) and with Video (represented as B)

Unimodal audio and video

- Accuracies of audio are significantly higher than the video in both cases
- Video accuracy in the unseen singer is close to chance indicating gestures are highly singer dependant

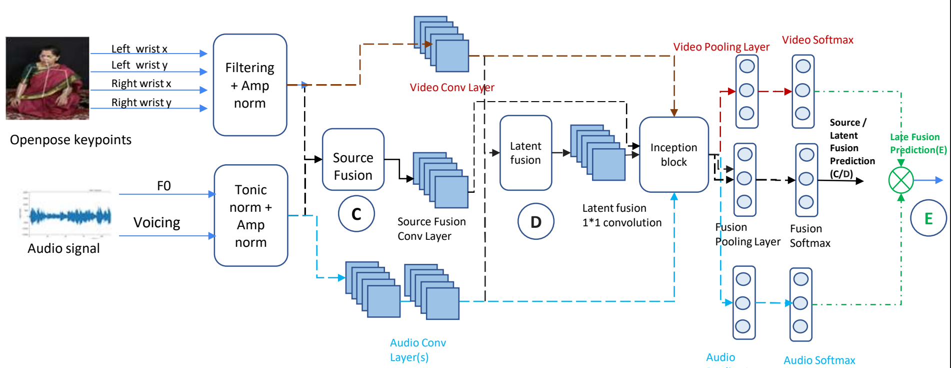
Data Split	Seen Singer		Unseen Singer	
	Audio	Video	Audio	Video
AG	92.1	36.3	76.9	14.3
CC	79.4	31.8	60.4	13.8
SCh	77.0	39.2	67.2	10.0

Table 4. Validation accuracy (%) of only audio and only video modalities on seen/unseen singer train-val splits.

Experiment 2

Multimodal raga prediction

Experiments (multimodal raga prediction)



Representation of multimodal predictions: source fusion (C), latent fusion (D), late fusion (E)

Results

- Source fusion performs very poorly ('overwhelmed' by the weaker modality)
- Latent fusion likely benefits from the inter-relations of the synced multimodal streams

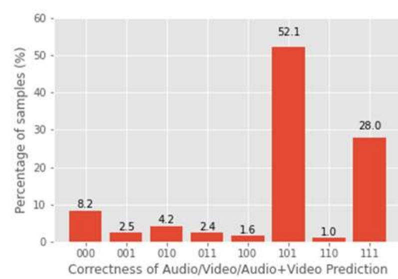
Model Type	Model Name	AG	CC	SCh	Mean
A	Video	36.3	31.8	39.2	35.8
B	Audio	92.1	79.4	77.0	82.8
C	Source fusion	30.1	42.4	35.8	36.1
D	Latent fusion	93.3	82.7	79.2	85.1
E1	Equal voting	85.9	73.7	67.9	75.8
E2	Stacking classifier – RF	81.9	74.2	76.3	77.5

Table 5. Validation accuracy (%) from each singer's split for different model architectures in the seen singer task.

Results

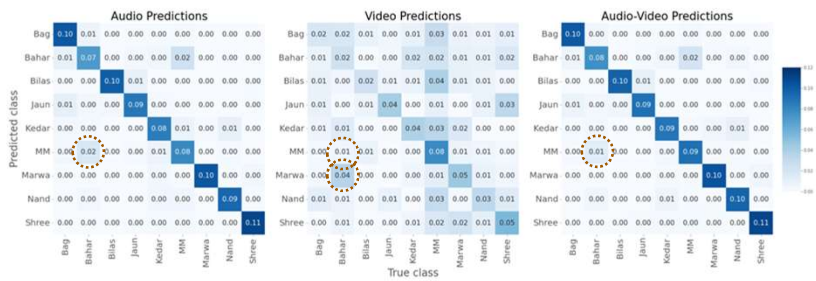
- Most common is to be predicted correctly by audio but incorrectly by video
- 6.6% video correct but audio incorrect
- 2.5% (75/3021 clips) correctly predicted by multimodal classification and incorrectly predicted by unimodal.
- Error analysis: Silence or rest; pitch errors

Video example: AG_4b_Nand: 0, 1:30, 2:40
https://www.youtube.com/watch?v=PhbAfXGJZLg&t=2s



Histogram indicating the number of correct predictions made by different modalities

Results



Conclusion

- We showed that the combination of <u>coordinated</u> audio and video features can enhance raga classification within short excerpts of a performance.
- Melodic improvisation in Indian classical music is viewed as motion in the “space” of a given raga. So, potentially, gestures can be mapped to entire sung phrases.
- The gestures are thought to have a nuanced and context-dependent relationship with melody, and a promising subject for deep learning…

References

- Martin Clayton, Preeti Rao, Nithya Shikharpur, Sujoy Roychowdhury and Jin Li, "Raga Classification From Vocal Performances Using Multimodal Analysis", Proceedings of ISMIR 2022

- M. Clayton, K. Jakubowski, and T. Eerola, “Interpersonal entrainment in indian instrumental music performance: Synchronization and movement coordination relate to tempo, dynamics, metrical and cadential structure,” Musicae Scientiae, vol. 23, no. 3, pp. 304–331, 2019.

- Rahaim, M., 2013. Musicking bodies: Gesture and voice in Hindustani music. Wesleyan University Press.

- Pearson, Lara (2013) 'Gesture and the sonic event in Karnatak music.', Empirical musicology review., 8 (1).

Thank you