The Journal of Acoustical Society of India

Volume 48 Number 1-2 January-April 2021 pp 1-98

ISSN 0973-3302

THE JOURNAL OF ACOUSTICAL SOCIETY OF INDIA

Volume 48 Number 1-2 January-April 2021

ACOUSTICAL SOCIETY OF INDIA

1971

॥ नादाधीनमतोजगत् ॥

A Quarterly Publication of the ASI

https://acoustics.org.in

The Journal of Acoustical Society of India

The Refereed Journal of the Acoustical Society of India (JASI)

CHIEF EDITOR:
B. Chakraborty
CSIR-National Institute of Oceanography
Dona Paula,
Goa-403 004
Tel: +91.832.2450.318
Fax: +91.832.2450.602
E-mail: bishwajit@nio.org

ASSOCIATE SCIENTIFIC EDITOR:

A R Mohanty
Mechanical Engg. Department
Indian Institute of Technology
Kharagpur-721302, India
Tel. : +91-3222-282944
E-mail : amohantyemech.iitkgp.ernet.in

Editorial Office:
MANAGING EDITOR
Mahavir Singh

ASSISTANT EDITORS:

Yudhisther Kumar
Devraj Singh
Kirti Soni
ASI Secretariat,
C/o Acoustics and Vibration Metrology
CSIR-National Physical Laboratory
Dr. KS Krishnan Road
New Delhi 110 012
Tel: +91.11. 4560.8317
Fax: +91.11.4560.9310
E-mail: asisecretariat.india@gmail.com

The Journal of Acoustical Society of India is a refereed journal of the Acoustical Society of India **(ASI)**. The **ASI** is a non-profit national society founded in 31st July, 1971. The primary objective of the society is to advance the science of acoustics by creating an organization that is responsive to the needs of scientists and engineers concerned with acoustics problems all around the world.

Manuscripts of articles, technical notes and letter to the editor should be submitted to the Chief Editor. Copies of articles on specific topics listed above should also be submitted to the respective Associate Scientific Editor. Manuscripts are refereed by at least two referees and are reviewed by Publication Committee (all editors) before acceptance. On acceptance, revised articles with the text and figures scanned as separate files on a diskette should be submitted to the Editor by express mail. Manuscripts of articles must be prepared in strict accordance with the author instructions.

All information concerning subscription, new books, journals, conferences, etc. should be submitted to Chief Editor:

B. Chakraborty, CSIR - National Institute of Oceanography, Dona Paula, Goa-403 004,
Tel: +91.832.2450.318, Fax: +91.832.2450.602, e-mail: bishwajit@nio.org

Annual subscription price including mail postage is Rs. 2500/= for institutions, companies and libraries and Rs. 2500/= for individuals who are not **ASI** members. The Journal of Acoustical Society of India will be sent to **ASI** members free of any extra charge. Requests for specimen copies and claims for missing issues as well as address changes should be sent to the Editorial Office:

ASI Secretariat, C/o Acoustics and Vibration Metrology, CSIR-National Physical Laboratory, Dr. KS Krishnan Road, New Delhi 110 012, Tel: +91.11.4560.8317, Fax: +91.11.4560.9310, e-mail: asisecretariat.india@gmail.com

The journal and all articles and illustrations published herein are protected by copyright. No part of this journal may be translated, reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission of the publisher.

Copyright © 2021, Acoustical Society of India
ISSN 0973-3302

Printed at Alpha Printers, WZ-35/C, Naraina, Near Ring Road, New Delhi-110028 Tel.: 9810804196. JASI is sent to ASI members free of charge.

B. CHAKRABORTY
Chief Editor

MAHAVIR SINGH
Managing Editor

A R MOHANTY
Associate Scientific Editor

Yudhishter Kumar Yadav
Devraj Singh
Kirti Soni
Assistant Editors

EDITORIAL BOARD

M L Munjal
IISc Banglore, India

Michael Vorländer
ITA Aachen, Germany

S Narayanan
IIT Chennai, India

V R SINGH
PDM EI New Delhi-NCR, India

R J M Craik
HWU Edinburg, UK

Trevor R T Nightingle
NRC Ottawa, Canada

N Tandon
IIT Delhi, India

J H Rindel
Odeon A/S, Denmark

G V Anand
IISC Banglore, India

Gopu R. Potty
University of Rhode Island, USA

S S Agrawal
KIIT Gurgaon, India

Yukio Kagawa
NU Chiba, Japan

D D Ebenezer
NPOL Kochi, India

Sonoko Kuwano
OU Osaka, Japan

Mahavir Singh
CSIR-NPL, New Delhi, India

A R Mohanty
IIT Kharagpur, India

Manell E Zakharia
ENSAM Paris, France

Arun Kumar
IIT Delhi, India

Ajish K Abraham
IISH Mysore, India

S V Ranganayakulu
GNI Hyderabad, India

The Journal of Acoustical Society of India

A quarterly publication of the Acoustical Society of India

Volume 48, Number 1-2, January-April 2021

ARTICLES

INFORMATION

FOREWORD

Editorial on Special Issue of "Musical Acoustics"

100 years ago, Sir C.V. Raman published a seminal paper on Musical Acoustics. It is a privilege for us to celebrate this by way of publishing several papers in frontiers of Musical Acoustics in this Special Issue of the Journal of Acoustical Society of India. 100 years is a pretty long time in history of Science whence we have seen tremendous advancement in every domain of science, precisely from macro to femto level. In India, we have nurtured musical acoustics since the 1920s - of course tools of theory and experiment changed and evolved with time. This Special Issue covers a wide range of musical acoustics with exhaustive studies and use of novel techniques. We hope this issue will not only be a piece of fantastic document to the community of musical acoustics, its interdisciplinary nature will fascinate young scientists, students, music practitioners and music therapists alike, to enrich musical acoustics with their knowledge in the domain they are working.

We take pleasure in presenting eleven (11) papers covering area from archaeo-acoustics to application of neural network in acoustics. In the first paper of this issue "Sensing the sound of Past: Acoustic Evaluation of Udaigiri Caves near Vidisha (Madhya Pradesh, India)", an interesting work has been reported where a study on acoustic evaluation of Udaygiri Caves has been performed with the help of parameters like reverberation time, early decay etc based on Impulse Response Recording. The study concludes for the first time that the cave 19 possesses most complex nature in regard to reverberation time contrary to earlier mentions.

The application of Neural Network architectures as an upcoming vista is so widespread that it has found its application even in hard core particle physics also. The domain of the second paper "Neural Network architectures to classify emotions in Indian Classical Music" is this neural network architecture applied to Indian Classical Music to classify emotions for the very first time. The field of deep learning is exhaustively used in different domain but this paper; perhaps the authors present a new dataset which has 400 audio clips, where 200 clips correspond to a happy emotion and remaining 200 to sad emotions. This technique is very rich for application in Indian Classical Music having wide range of emotions difficult to classify rigorously with usual techniques.

Sir C.V. Raman in his famous paper published in Nature on *Tabla* and *Mrudangam* reported "Five such harmonics (inclusive of the fundamental tone) can be elicited from the drumhead in this type of instrument, the first, second, and third harmonics being specially well sustained in intensity and giving a fine musical effect". In the third paper of this issue "Automatic Stroke Classification of Tabla Accompaniment in Hindustani Vocal Concert Audio", authors present different acoustic features that capture the distinctive characteristics of tabla strokes and envisage developing an automated system which predict the label as one of a reduced, but musicologically motivated, target set of four stroke categories. This study looks to build an instrument-independent stroke classification system for accompaniment table based on available tabla solo audio tracks.

In Hindustani music, a *"Gharana"* or school refers to the adherence of a group of musicians to a particular musical style. The fourth paper "Variation of singing styles within a particular Gharana of Hindustani classical music - a nonlinear multifractal study" aims to study quantitatively the evolution of singing style among four artists of four consecutive generations from *Patiala Gharana* using non linear multifractal analysis (MFDFA) technique. The observations from the variation of spectral width look to scientifically evaluate a philosophical term - *Guru-Shisya Parampara* (teacher-student tradition) in Indian Classical Music.

In Hindustani (North Indian) classical music, the most common way to classify a raga is under ten parent scales (called *thaat*). A *thaat* is no more than a seven-note scale including one each of the seven notes *sa re ga ma pa dha ni* (the Indian equivalents of *do re mi fa so la ti*). The paper "Probabilistic Analysis of Three Ragas of the Bhairavi Thaat" attempts to provide a mathematical approach for different ragas of the same *thaat* by analysing their probabilities. The analysis covers the conditional and unconditional probabilistic behaviour of the three different ragas namely *Raga Bhairavi, Raga Bilaskhani Todi* and *Raga Malkauns*. The study includes the comparative study of aaroh-avroh (ascent-descent sequential pattern of the three ragas as well as comparison of the strings (note assemblies) used in the *bandishes*.

Rhythm and tempo are two of the most important features of Indian Classical Music which evokes and modify emotional appraisal. It has been shown that features of a *tāla* - like beats, vocables, structure, cyclical beat-structure, metre *etc.* contribute to the total impact on the perception of music. In the following paper "The Perception of Features and Emotions in the *Rāgas* When They Are Composed in Different Patterns of Beats, Vocables and Rhythms", the authors conducted a survey on four *rāgas* and four *tālas* - *Khamāja, Mānda, Mālkaunsa* and *Darbāri Kānadā* using *tālas, Teentāla, Dādarā, Ektāla* and *Choutāla*. It was found that when perception is rated as positive, it is correlated with excitement and perceived complexity. On the other hand, negative perception leads to the selection of features such as less positive and slow.

In the next paper "Madhyama Rahitha "Melakartha" Scales in Carnatic Music" the author proposes a new scheme of 156 melakartha scales in Carnatic music. including the Venkatamakhi scales, vikrutha panchama scales and the 48 madhyama rahitha scales. This has applications in some of the derived (Janya) ragas (consisting of less than 7 independent notes taken from any melakartha scale) which are common for more than one melakartha scale and result in innumerable raga system.

In Hindustani Music (HM), pause or silence evoke the essence in a raga that controls the finest emotional behavior. In the following paper "Identifying Style of Vocalist using Silence in Hindustani Music", the authors have measured the silence / pause parts of eighty-six musical signals sung by twenty-two eminent vocalists when each of them sung same four ragas. This work tries to to understand the singing style in the light of pauses using features like note sequences (phrase), total number of pauses and time duration of both pause and phrases, sequences of phrase arrangement at the onset and offset etc. This technique of using silence to quantify the style of an artistic performance is unique even in global scenario.

The human brain analyses the complex auditory environment, perceptually grouping the sounds arriving from one source and separates them from competing sources by a phenomenon is called auditory stream segregation. In the paper "Auditory stream segregation in singers and non-singers: A comparative study", the authors look forward to evaluate and compare auditory stream segregation abilities across singers and non-singers using spectral profile of the respondents. The survey was done on 30 singers (with Carnatic music training for at least three years) and 30 non-singers and the findings of this study showed that the spectral profiling abilities of singers were not significantly different from the non-singers group.

In Indian aesthetic tradition - *alamkara* (lit. ornamentation) or embellishment is considered a key aesthetic component that 'embellishes' or beautifies a piece. In the paper "Perception of Ornamentation in Hindustani Classical Music", the authors address the issue whether *alamkaras* merely beautify a piece or do they also perform specific functions in terms of emotional perception of music and other expressive features. 105 participants listened and rated four raga clips, two happy and two sad, rendered using three *alamkaras* - *meand, murki,* and *khatka* in their non-articulated forms. The results suggest that ornamentations play a key role, not only in the intensification of primary emotions but also in variations in secondary associated emotions, which change the entire color of the raga rendition. The findings of this work are relevant in the context of Indian musicology as they throw new light on the specific roles key *alamkaras* play in coloring emotions and associated expressive features

The concluding paper of this issue "*Ragas* in Bollywood music - A microscopic view through multrifractal cross-correlation method" deals with an interesting topic where the authors look to explore particular features of a certain *raga* which make it understandable to common people and enrich the song in which that raga is used.For this two common ragas, *"Bhairav"* and *"Mian ki Malhar"* were chosen and multifractal cross correlation is evaluated between parts of raga and film songs in which they are used. A scientific basis of the amount of correlation that exists between the raga and the same raga being utilized in Film music has been established. This will help in generating an automated algorithm through which a naïve listener will relish the flavor of a particular raga in a popular film song.

In the end, we take the opportunity to thank Editor-in-Chief Dr. Biswajit Chakraborty and the Editorial Board of JASI to give our team the privilege to serve as Editor for this special issue on a very interesting and intriguing topic very much relevant for the contemporary readers. We would also like to thank the reviewers who provided valuable and timely feedback on the manuscripts. Looking forward to the future, it would not be impertinent to say that an interesting work in the field of musical acoustics would be to revisit Sir C.V. Raman's outstanding research in regard to higher harmonics of Indian *tabla* and *mrudangam*, using frontier acoustical analysis techniques developed over the years. The tradition of Indian Music is like an ocean, we have been able to touch upon only a few of its waves in this Special Issue, and the entire ocean is left wide open for future scholars and researchers to explore. It was a real pleasure for us to serve as Guest Editors of this Special Issue on a topic which is very close to our brain and soul; we sincerely hope that this Special Issue of interdisciplinary nature will evoke interest in present scholars, students and interested readers of musical acoustics and our efforts would be fruitful if more sincere researchers of science come forward to explore this nascent field.

Prof. Dipak Ghosh
Dr. Ranjan Sengupta
Dr. Shankha Sanyal
Dr. Archi Banerjee
Guest Editor - JASI
Sir C.V. Raman Centre for Physics and Music
Jadavpur University, Kolkata-700032, India

The Journal of Acoustical Society of India : Vol. 48, No. 1-2, 2021 (pp. 1-10)

Sensing the sound of Past: Acoustic Evaluation of Udaigiri Caves near Vidisha (Madhya Pradesh, India)

Ajinkya S. Umbarkar[1], Deoram V. Nandanwar[1] and Omprakash P. Chimankar[2]
[1]Department of Physics, SMM College of Science, Nagpur-440009, India
[2]Department of Physics, RTM Nagpur University Campus, Nagpur-440033, India
e-mail: ajinkya22007@gmail.com

[Received: 20-11-2020; Revised: 29-12-2020; Accepted: 29-12-2020]

ABSTRACT

Here we report first ever study on acoustic evaluation of Udaigiri Caves near Vidisha (Madhya Pradesh, India). In this study we used ambisonic recorder. We then studied room acoustic parameters like Reverberation Time, Early Decay Time, Clarity, Definition etc., based on the Impulse response recording. FFT analysis of each measurement gives many resonance peaks, hence we also suggest further study to explore effect of those musical notes on the brain. Keeping in mind of the presence of musical sculpture on the lintel of Cave 4, it will be interesting to do further research by studying it's response to the similar kind of instrument played inside. We finally conclude that Cave 19 at this complex is most reverberant contrary to earlier mentions, not the experimental ones. Also there is need to study this caves for the response to their respective resonance frequencies which holds the importance in archaeoacoustical studies.

1. INTRODUCTION

Archaeoacoustic as the name suggest is an amalgamation of techniques in Acoustics and the historical knowledge from archaeology / history. In Archaeoacoustics, acoustics of a place under investigation is studied using different techniques[1].

In India, archaeoacoustical investigation started in 1990s-2000s. Many sites like Hulimavu Cave Temple (Bengaluru, Karnataka), Udayagiri Cave (Odisha), Koothambalam of Vadakkunathan Temple (Trissur, Kerala) *etc.*, have been studied by some researchers[2]. This area has seen growing importance and attention in the last decade. These investigations were done with normal recording microphones or handheld recorders of Sony, Zoom *etc*. Some years ago Umashankar Manthrawadi developed his own First Order Ambisonic recorder named *"Brahma"*, with the help of this he carried out his work at many sites like Anupu (Andhra Pradesh), Udayagiri (Odisha) and some Koothambalam of various Kerala temples[3]. First author personally carried out survey and measurement at Bhojeshwar Temple (Bhojpur), Bhimbetka.

2. UDAIGIRI CAVES

It is a group of 20 caves located [Coordinates: 23°32′11.0′N 77°46′20′E] near Vidisha (Madhya Pradesh). They lie 6 Km West to Vidisha, 11 Km Northeast of Sanchi, and 60 Km Northeast of Bhopal. They were carved at the end of 4th century CE to early part of 5th century CE. Site was first studied in detail by Sir

© 2021 Acoustical Society of India

Alexander Cunningham in 1870s and were discussed in his *"Reports of Tours in Bundelkhand and Malwa in 1874 -75 and 1876 -77"* published by Archaeological Survey of India as Volume 10 of the reports. Caves 1, 3 to 6 and 13 have rich sculptural contents. There are 19 Hindu and 1 Jain caves[4].

During the visit we interacted with local people and security. We discussed about the site. Based on the inputs from them Cave 4, 6, 19 were selected for further experimental work.

2.1 Description of the Caves

Cave 4 : Sir Alexander Cunningham named it as *"Vina Cave"* because of carving of Vina (harp) player on its lintel. An Ekamukhi (with one face) Shivalinga is present. Cave measures 15 ft. ×12 ft.[5].

Cave 6: It has inscription of Gupta Year (GE) 82 (= 401-402 CE) on its lintel just above the entrance. This inscription tells us that Gupta King Chandragupta 2 and his minister Virasena visited this caves. Cave measures 15 ft. × 12 ft. This cave is sometimes referred as *"Chandragupta's Cave"*[6].

Cave 19 : Known as *"Amrita Cave"* it measures 22 ft. × 19.33 ft. An inscription in Nagari script dated to 1037 CE states that - Kanha, a pilgrim donated some resources to this temple. It also mentions that this temple was made on the orders of Chandragupta. Sculpture of *Samudra Manthana (Churning of Ocean)* scene is carved on its lintel and therefore Cunningham called is as *Amrita Cave*[7].

Fig. 1. (a)- Ekamukhi Shivalinga in Cave 4 (Vina Cave), (b)- Cave 6, (c)- Cave 19 *(Amrita Cave)* at Udaigiri (Photos by Author).

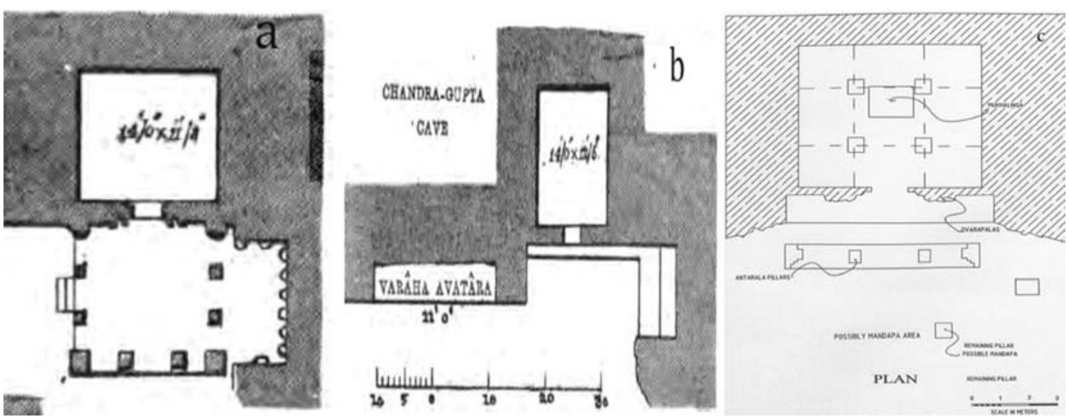


Fig. 2. (a)- Ground plan of Cave 4, (b)- Cave 6 (*Source:* Wikipedia) and (c)- (*Dass* 2001) Cave 19 of Udaigiri Caves.

3. METHODOLOGY

Our methodology involves following steps :-

- Interaction with local peoples, ASI officials (if site is protected by them) about how this places sounds, are there any earlier mentions of such phenomena and its records etc.
- Recording of Impulse response (IR) using balloon pop) and Zoom H3-VR (a FOA=First Order Ambisonic recorder). After recording processing for studying room acoustical parameters will be done using Audacity and Adobe Audition 3.
- For analysis of FFT (Fast Fourier Transform) Audacity will be used. Room Acoustics Parameters like C (Clarity), RT or T (Reverberation Time), EDT (Early Decay Time) etc will be studies using Aurora Plugin with Audacity. For STI (Speech Transmission Index) analysis STI function in Aurora Plugin with Adobe Audition will be used.
- Aurora Plugins, based on ISO: 3382 parameters was developed by Prof Angelo Farina.
- Zoom Ambisonics Player for converting A-format (raw) to B-format (post processed). This is needed if one wants to study the spatial parameters - from where and how much energy is coming.

3.1 Room Acoustic Parameters discussed

EDT : It is based on 0 to 10 dB of initial decay. Influenced by early reflections, it depends upon measuring position of source-receiver and room geometry [9]. This parameter is measured from a linear regression line fit to the initial portion of the impulse response and computed separately for different frequency of octave band[10].

$$\mathrm{EDT}=\mathrm{RT}_{0,-10} \text{ (in Sec)} \tag{1}$$

D_{50} : It is evaluated at 500 Hz, 1 KHz, 2 KHz, 4 KHz and should have values >50% for good speech intelligibility[11]. D will be 100% if the impulse response does not contain any components with delays in excess of 50 ms[12].

$$D_{50} = \int_0^{50} h^2(t)\,dt \,/ \int_0^{50} h^2(t)\,dt\,(in\ dB) \tag{2}$$

Where $h^2(t)_{0\text{-}50}$ms = RIR measured at some distance preferably 1m, $h^2(t)_{0\text{-}\infty}$ is a RIR measured with omni directional source.

C_{80} : It is a measure of transparency of musical structures[13]. In practice an unweight average of 125 Hz - 4 KHz is used[14]. Optimal value for C80 are -4 to 0 dB (orchestral music), 0 to 4dB (singers), -4 to 4 dB (general purpose)[15]. By substituting 0 to 50 limit in eqn 3 as 0 to 80 this parameter can be obtained.

C_{50} : It is based on the Haas Effect for speech *i.e.* when an acoustical reflection reaches within 50 ms of the direct sound. It improves when strong early reflections are present in a room. Optimal values for C_{50} are $\geq$3 dB[16]. A weighted average for 500 Hz to 4 KHz is often used for assessment of a room[17].

$$C_{50} = 10\ \log\left(\int_0^{50} h^2(t)\,dt \,/ \int_{50}^{50} h^2(t)\,dt\right)(in\ dB) \tag{3}$$

Where *h(t)* = power in first 50 ms and thereafter.

Reverberation Time (RT) : According to ISO 3382 measured RT is obtained by extrapolation of a 60 dB line fitted to a decay curve. In reality, it is difficult to achieve SNR which allows 60 dB decay range, therefore -5 to -35 dB decay is used *i.e.* T_{30}[18]. T_{10} is calculated for range -5 to -15 db of decay. T_{20} is an estimation of the 60 dB decay time by extrapolation. SNR of minimum 35 dB is preferred for measuring T_{20}[19]. For greater speech intelligibility RT should be in range of 0.8-1.0s[20]. It is determined by Sabine equation as follows:

$$T_{60} = 0.1611\times(V/S\alpha) \tag{4}$$

Where V = Volume of room in m^3, S = Total surface area in m^2, α = absorption coefficient

4. RESULTS AND DISCUSSION

We first carried out FFT analysis using Audacity. Results are as shown in figure 3. Here x-axis denotes frequency (in Hz) and y-axis is of SPL (in dB).

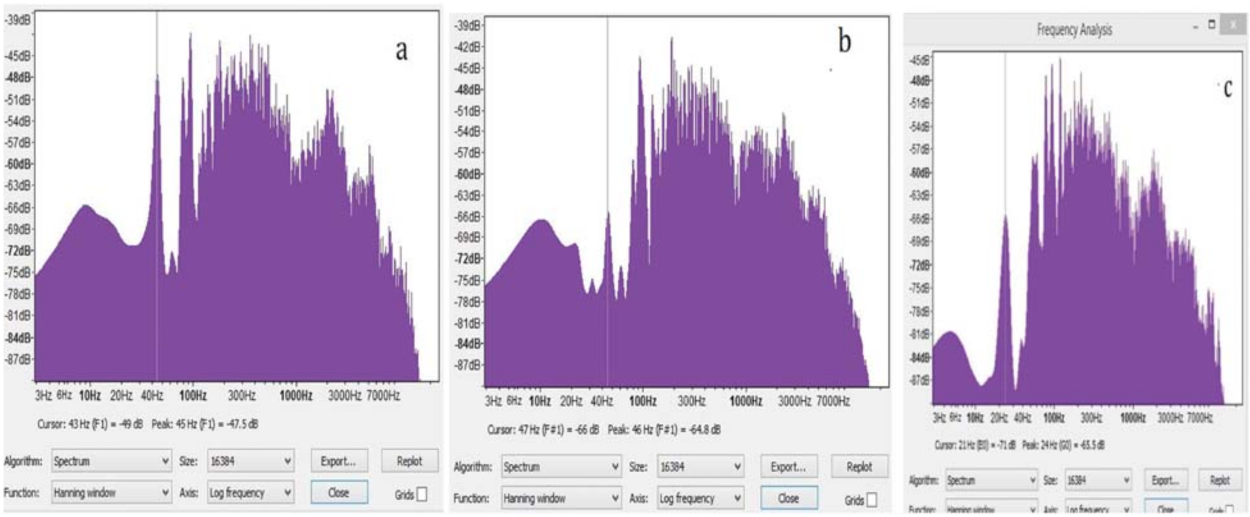


Fig. 3. (a)- Cave 4, (b)- Cave 6, (c)- Cave 19 of Udaigiri Caves.

Shows many resonance peaks on low frequency side. Table 1 lists low frequency, lowest SPL and musical note corresponding to it.

Tables 2, 3, 4 shows room acoustic parameters like EDT, T30, C50, D50. Here CH= Channel number.

Table 1. FFT peaks (Frequency with corresponding dB level and musical note).

Sr No	Cave	Frequency (Hz)	SPL (dB)	Musical Note
1	Udaigiri Cave 4	9, 45, 80, 94, 357	65.6, 47.5, 48.1, 41.7, 39.3	C#-1, F#1, D2, F#2, F4
2	Udaigiri Cave 6	10, 46, 93, 187	66.2, 64.8, 42.8, 38.2	E-1, F#1, F#2, F#3
3	Udaigiri Cave 19	24, 60, 78, 93, 120, 218	73.9, 66.8, 58.9, 57.7, 56.3, 56.6	G0, B1, D#2, F#2, A#2, A3

Table 2. Udaigiri Cave 4.

EDT	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	13.81	6.743	4.19	2.84	2.08	1.36	0.91	0.75	0.55	0.371
	CH2	10.78	5.063	4.45	2.88	1.91	1.37	0.86	0.70	0.54	0.354
	CH3	13.75	10.55	5.05	2.69	2.24	1.29	0.92	0.76	0.55	0.387
	CH4	12.53	5.558	3.92	3.06	1.96	1.41	0.90	0.75	0.55	0.377
T30	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1		35.136		3.784	2.194	1.45	0.953	0.759	0.561	0.398
	CH2	16.626	35.122	7.833	3.828	2.108	1.413	0.953	0.746	0.555	0.39
	CH3	16.088	34.518	8.78	4.194	2.236	1.476	0.946	0.761	0.571	0.396
	CH4	8.019	35.224		3.87	2.181	1.463	0.943	0.761	0.563	0.393

Conted.....

C50	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	-9.322	-5.563	-7.07	-5.17	-6.28	-4.25	0.455	2.469	4.125	6.893
	CH2	-1.66	-6.877	-8.88	-5.67	-3.51	-3.02	1.61	2.603	3.726	7.975
	CH3	-9.314	-7.071	-9.46	-6.18	-7.27	-4.46	-0.13	1.721	3.486	6.842
	CH4	-7.565	-5.656	-6.97	-5.16	-3.93	-4.09	-0.59	2.139	3.409	6.97
D50	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	10.466	21.741	16.41	23.32	19.05	27.34	52.61	63.84	72.11	83.02
	CH2	40.56	17.031	11.46	21.31	30.83	33.29	59.16	64.55	70.22	86.25
	CH3	10.484	16.407	10.17	19.41	15.78	26.37	49.26	59.78	69.06	82.86
	CH4	14.906	21.379	16.73	23.37	28.78	28.05	46.63	62.07	68.67	83.27
C80	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	-7.392	-4.468	-5.58	-2.12	-2.71	-0.80	3.539	5.578	8.219	12.86
	CH2	-0.663	-5.917	-7.31	-3.13	-0.76	-0.45	4.389	5.49	7.829	12.67
	CH3	-7.618	-6.144	-7.06	-3.98	-3.66	-1.14	2.583	4.8	7.611	12.14
	CH4	-5.803	-4.501	-4.48	-2.25	-0.81	-1.46	3.205	5.362	7.689	11.74

Table 3. Udaigiri Cave 6.

EDT	**Frq.band [Hz]**	**31.5**	**63**	**125**	**250**	**500**	**1k**	**2k**	**4k**	**8k**	**16k**
	CH1	8.594	8.032	6.443	2.993	1.946	1.298	0.973	0.8	0.539	0.356
	CH2	22.794	9.533	5.348	3.11	2.103	1.276	0.973	0.839	0.614	0.424
	CH3	18.51	8.844	5.674	3.001	1.967	1.351	0.961	0.77	0.588	0.391
	CH4	17.891	8.504	6.074	3.064	1.933	1.406	1.025	0.863	0.617	0.419
T30	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1			12.76	4.542	2.175	1.443	0.999	0.865	0.59	0.413
	CH2	32.978	16.45	9.172	4.362	2.155	1.406	0.992	0.876	0.584	0.427
	CH3	18.073	16.33	11.53	4.536	2.118	1.371	1.024	0.955	0.59	0.421
	CH4		24.04	9.233	4.476	2.133	1.41	1.021	0.944	0.594	0.414
C50	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	3.544	-1.76	-6.49	-2.98	-2.66	-0.49	0.286	1.565	4.207	7.6
	CH2	-10.64	-15.89	-8.83	-5.83	-6.48	-1.89	-0.41	0.884	3.04	6.045
	CH3	-2.791	-9.13	-6.85	-5.50	-3.10	-1.57	0.965	1.401	3.554	6.753
	CH4	-5.511	-5.97	-7.98	-5.99	-4.52	-2.40	-1.9	0.284	2.848	5.195
D50	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	69.337	39.98	18.33	33.47	35.17	47.19	51.65	58.91	72.48	85.19
	CH2	7.943	2.51	11.57	20.71	18.37	39.31	47.60	55.07	66.82	80.09
	CH3	34.467	10.89	17.12	21.97	32.86	41.05	55.53	57.99	69.39	82.56
	CH4	21.944	20.19	13.72	20.11	26.09	36.51	39.23	51.63	65.83	76.78
C80	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	4.445	-1.29	-4.14	-1.57	-0.08	2.105	3.13	4.7	8.426	12.88
	CH2	-9.576	-9.09	-6.01	-3.07	-3.43	1.474	2.695	4.024	6.905	10.91
	CH3	-2.65	-5.57	-5.15	-2.95	-0.45	1.584	4.05	4.915	7.457	11.49
	CH4	-3.634	-5.29	-7.29	-2.74	-1.42	0.168	2.139	4.033	6.887	10.83

Table 4. Udaigiri Cave 19.

EDT	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	38.66	21.87	11.98	4.58	2.62	1.58	1.43	0.99	0.60	0.39
	CH2	37.01	21.27	12.88	4.77	2.58	1.56	1.37	0.93	0.62	0.33
	CH3	38.95	21.44	13.68	4.22	2.46	1.67	1.39	1.07	0.68	0.42
	CH4	41.96	21.54	12.47	4.53	2.43	1.64	1.36	0.97	0.56	0.38
T30	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	24.086	20.636	15.653	5.54	2.873	1.914	1.437	1.064	0.719	0.466
	CH2	24.812	18.931	15.921	5.941	2.89	1.874	1.444	1.032	0.708	0.44
	CH3	25.317	19.48	16.226	5.576	2.904	1.865	1.431	1.032	0.72	0.451
	CH4	27.858	20.737	15.825	5.658	2.943	1.849	1.402	1.028	0.722	0.465
C50	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	-8.728	12.959	-11.063	-7.325	-5.401	-0.943	-1.517	0.37	3.804	7.37
	CH2	-8.144	-10.787	-11.256	-6.979	-3.068	0.048	-0.534	0.831	3.547	8.692
	CH3	-11.57	-12.307	-9.606	-6.278	-4.616	-3.623	-2.52	-1.332	3.131	6.633
	CH4	-14.703	-10.5	-9.363	-5.173	-2.236	-1.285	-0.83	0.481	4.297	7.42
D50	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	11.82	4.816	7.26	15.62	22.38	44.59	41.36	52.13	70.59	84.51
	CH2	13.293	7.701	6.967	16.70	33.04	50.27	46.93	54.77	69.35	88.09
	CH3	6.512	5.552	9.868	19.07	25.67	30.27	35.89	42.39	67.28	82.16
	CH4	3.275	8.183	10.377	23.31	37.40	42.66	45.23	52.77	72.89	84.67
C80	Frq.band [Hz]	31.5	63	125	250	500	1k	2k	4k	8k	16k
	CH1	-8.287	-9.555	-8.577	-5.248	-2.718	1.187	1.729	4.107	7.556	12.01
	CH2	-7.97	-8.475	-9.606	-3.529	-1.369	2.179	2.441	3.994	7.088	13.07
	CH3	-11.12	-8.761	-5.959	-3.795	-1.852	-0.665	0.539	2.399	6.506	11.15
	CH4	-13.103	-8.192	-6.04	-2.986	-0.621	0.562	2.04	3.604	7.941	11.75

Table 5. Room Acoustic Values in condense form.

Sr No	Site name	C50	C80	D50	EDT	T10	T20	T30	STI Male	STI Female
1	Cave 19	-3.45	-0.66	36.19	8.55	9.25	8.76	7.55	0.485	0.508
2	Cave 6	-1.98	0.99	41.44	4.07	5.65	7.04	6.61	0.536	0.556
3	Cave 4	-2.27	0.64	39.45	3.29	3.955	6.245	6.71	0.544	0.568

Above table gives the values in condensed form. For this we first took average of all channel values for each frequency. These values were then added and averaged for each octave band. Here we have values of Clarity (C50, C80), Definition (D50), Early Decay Time (EDT), Reverberation Time (T10, T20, T30), Speech Transmission Index (Male, Female).

Table 6. IACC Values.

Freq.[Hz]	31.5	63	125	250	500	1000	2000	4000	8000	16000
Cave 4	0.807	0.936	0.559	0	0.547	0.535	0.151	0.304	0.29	0.579
Cave 6	0	0.605	0.095	0.244	0.493	0.523	0.396	0.288	0.27	0.574
Cave 19	0.463	0.864	0.608	0.055	0.295	0.449	0.332	0.601	0.54	0.768

IACC value denotes the correlation between the two signals arriving at two ears. +1 denoted correlation with out of phase, 0 for no correlation, + perfect correlation.

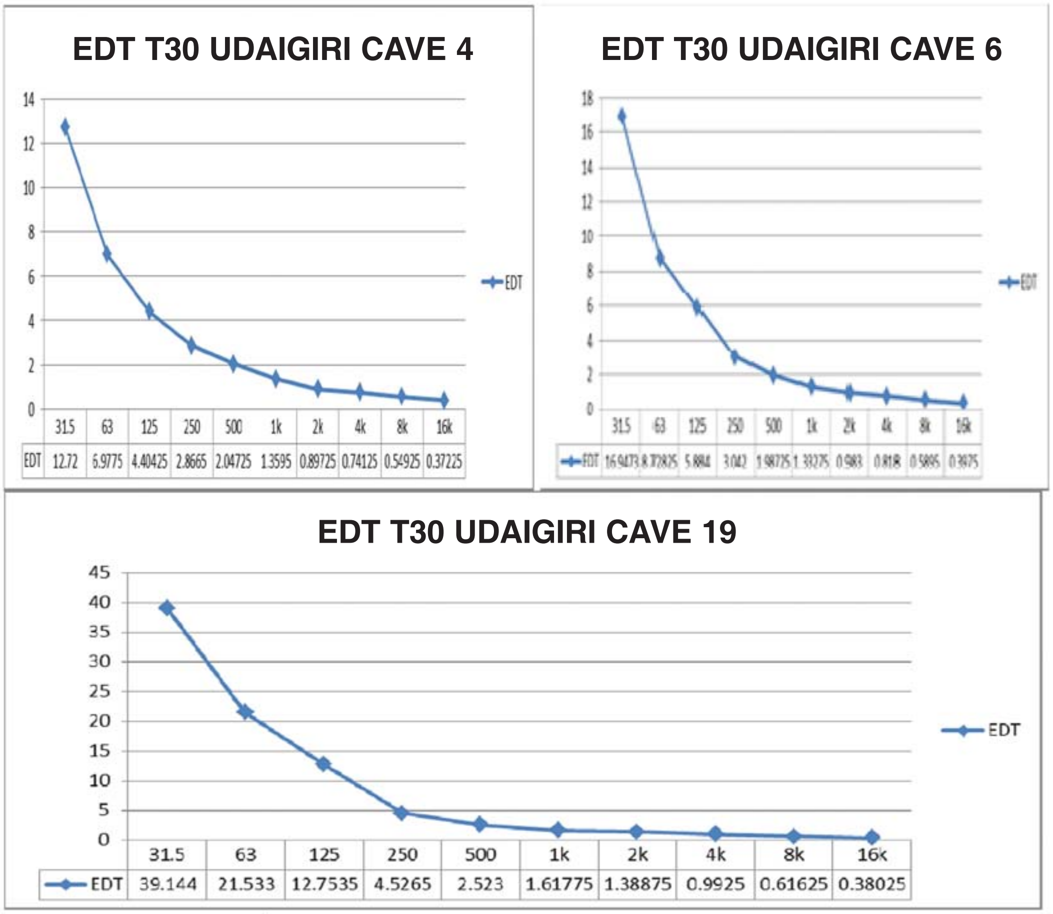


Fig. 4. EDT T30 values of Cave 4, 6, 19 of Udaigiri Caves.

Here x-axis denotes octave band frequencies (in Hz) and y-axis denotes time (in seconds)

5. DISCUSSION & CONCLUSION

It can be seen in Fig 4. that at low frequencies up to 125 Hz values of EDT are very high. Is it due to resonances or some other factors, is yet to be resolved. One possibility is that it is due to instrumental response at those frequencies. From 250 Hz and above the response shows a decreasing trend [personal communication with Daniel Courville]. Such behavior was also observed in Reverberation Time (T10, T20, and T30) curves. IACC values (in Table 6) show that all caves show low correlation values at 250 Hz and 2 kHz. Low correlation values of less than 1 are due to the reverberant environment of the caves. From the Table 5, it appears that Cave 19 shows high values for EDT as well as RT compared to other investigated caves at the same site. Due to this reason, for the same cave values of clarity, definition are low. High RT values in Cave 19 gives much more reverberation environment and thereby causes more pleasing sensation to ear.

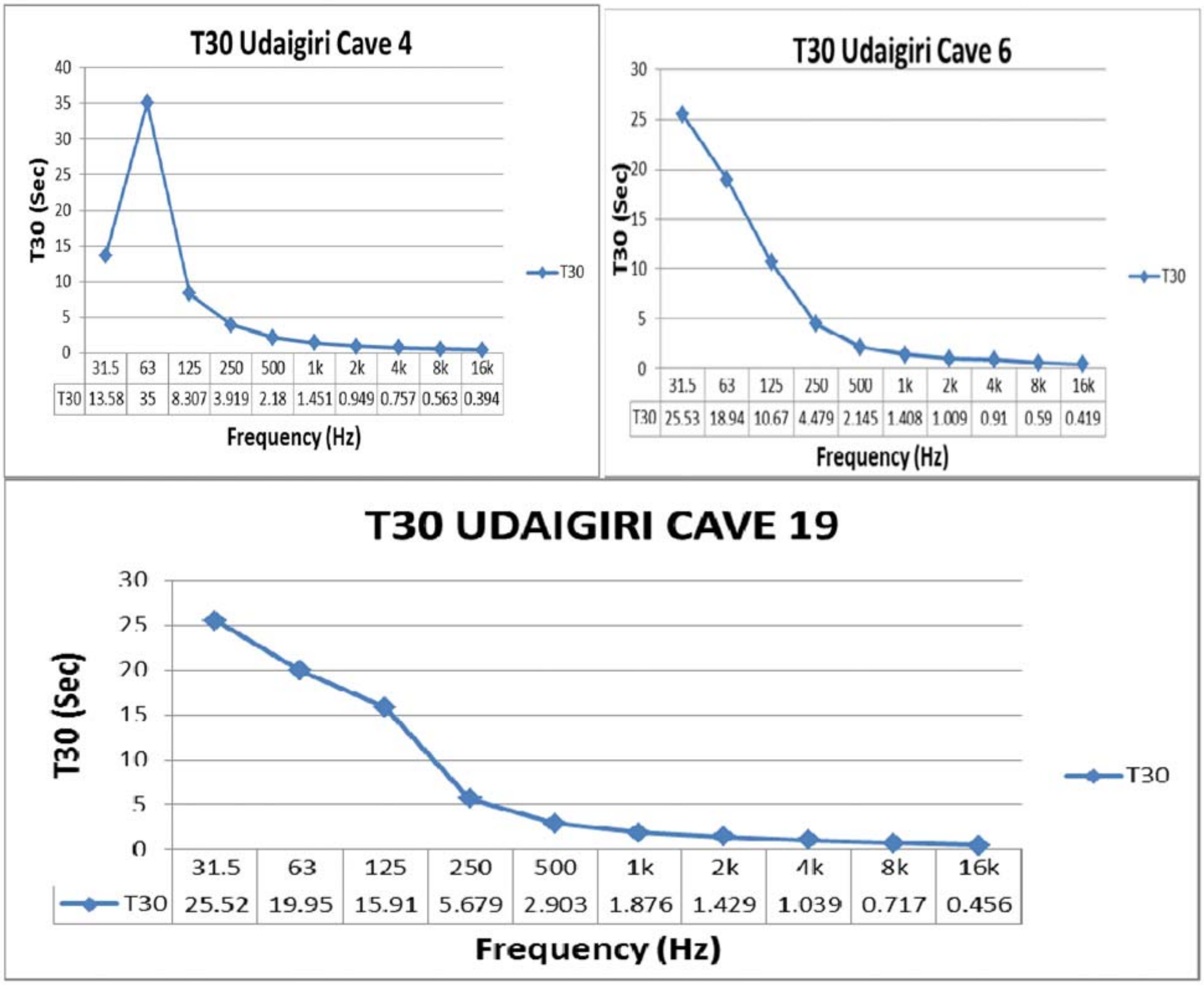


Fig. 5. T30 values of Cave 4, 6, 19 of Udaigiri Caves.

"Other than the acoustical quality of Cave 4 is also of significance. The presence of the musical instruments on the lintel of Cave 4 may be relevant when we notice the acoustical quality of the cave. The amplification of sound without distortion or echo creates vibrations even when slokas are recited at a low pitch. This quality is highly noticeable. The acoustical quality was found missing in the other caves on this site. Presence of scenes of music and dance in the Varaha Panel nearby, where the descent of Ganga is accompanied with celestial dance and music, can be seen as part of the overall scheme. The five musicians that accompany the descent of the river goddesses carry vina, lute and flute. Narada who is the inventor of vina is shown to the south of the Varaha, a distinctive position. Tumburu who is also a Gandharva musician, stands next to him playing a guitar"[21].

As a part of any structure, acoustic is present there. Dass (in the above passage) while discussing acoustical impression opines that, Cave 4 acoustic is best. In our study, we have found that it is not so. Infact Cave 19 acoustic is noticeable in all aspects (*i.e.* various acoustic parameters) based on the experimental data. STI values (Male / Female) in Table 5 indicates that speech transmission index falls in "Fair" category of STI and CIS scale proposed by Barnett in 1995.

Further there is need to work on the relation between Musical notes (indicated in Table 1) and their effect on the brain. We also suggest to check the magnetic field and other anomaly (if present) with proper instruments. This study is limited only for objective acoustic parameters, hence there is scope for work in subjective assessment (person related). We also suggest the use of TRV camera and magnetic field detector for future studies along with 3D reconstructions. Further there is a need to research on how the music instruments like Vina, Lute and Flute tuned to specific notes (Saptaswaras) as well as to the resonating frequency of all the three caves investigated responds. We also suggest that to carry out soundscape study in future.

Conflict of Interest : None

6. ACKNOWLEDGEMENT

We are thankful to Prof. S Rajagopalan (Nagpur), Prof. Angelo Farina (University of Parma, Italy), Dr. MA Tavares (Goa). We are thankful to all officials of ASI Bhopal Circle for their support in circle's office and at site.

7. REFERENCES

[1] http://en.wikipedia.org/wiki/Archaeoacoustics#:~:text=Archaeoacoustics%20is%20the%20use%20of%20acoustical%20study%20as,the%20study%20of%20the%20acoustics%20of%20archaeological%20artifacts. (accessed on 22.10.2020)

[2] A.S. Umbarkar, D.V. Nandanwar and O.P. Chimankar, 2018. "Archaeoacousticsal investigation of sacred & worship places : how to do it?", WESPAC 2018, New Delhi, India, November 11-15, 2018, pp. 1-6.

[3] https://brahmamic.com/ (accessed on 24.10.2020)

[4] https://en.wikipedia.org/wiki/Udayagiri_Caves (accessed on 26.10.2020)

[5] https://en.wikipedia.org/wiki/Udayagiri_Caves (accessed on 26.10.2020)

[6] https://en.wikipedia.org/wiki/Udayagiri_Caves (accessed on 26.10.2020)

[7] https://en.wikipedia.org/wiki/Udayagiri_Caves (accessed on 26.10.2020)

[8] M. I. Dass, 2001. "UDAYAGIRI: A SACRED HILL ITS ART, ARCHITECTURE AND LANDSCAPE", De Montfert University, Leicester, UK.

[9] M. D. Topa, N. Toma, B. S. Kirei, I. Sărăcut and A. Farina, 2012. Experimental acoustic evaluation of an auditorium. *Advances in Acoustics and Vibration,* Article ID 868247, https://doi.org/10.1155/2012/868247

[10] B. Rakerd, E. J. Hunter, M. Berardi and P. Bottalico, 2018. Assessing the Acoustic Characteristics of Rooms: A Tutorial With Examples. *Perspectives of the ASHA special interest groups*, 3(19), 8-24. doi:10.1044/persp3.SIG19.8.

[11] M. D Topa *et al.*, https://doi.org/10.1155/2012/868247

[12] H. Kuttruff, 2009. Room Acoustics. 5th Ed, Spon Press, p. 224.

[13] R. Till, 2017. An archaeoacoustic study of the Ħal Saflieni Hypogeum on Malta. *Antiquity*, **91**(355), 74-89.

[14] M. Kleiner, D. L. Klepper and R. R. Torres, 2014. Worship Space Acoustics. *J Ross Publishing, Second Indian Reprint*, p. 62.

[15] V. G. Gonzalez, 2009. Acoustic study and conditioning for Kirche Am Hohenzollernplatz in Berlin (Doctoral dissertation, Ph.D. Thesis, Institut Fur Technische Akustik). https://upcommons.upc.edu/bitstream/handle/2099.1/7582/Final%20Thesis.pdf;sequence=1

[16] B. Rakerd *et al.*, 2018. doi:10.1044/persp3.SIG19.8.

[17] M. Kleiner *et al.,* 2014. p. 62.

[18] https://www.acousticbulletin.com/room-acoustic-descriptors-rt-c50-and-gain

[19] B. Rakerd *et al.*, 2018. doi:10.1044/persp3.SIG19.8.

[20] M. Kleiner *et al.*, 2014. p. 22.

[21] M. I. Dass, 2001. pp. 165-166.

The Journal of Acoustical Society of India : Vol. 48, No. 1-2, 2021 (pp. 11-19)

Neural network architectures to classify emotions in indian classical music

Uddalok Sarkar[1*], Sayan Nag[2], Medha Basu[1], Archi Banerjee[1,3], Shankha Sanyal[1,4], Ranjan Sengupta[1] and Dipak Ghosh[1]

[1]*Sir C.V. Raman Centre for Physics and Music, Jadavpur University, India*

[2]*Department of Medical Biophysics, University of Toronto, Canada*

[3]*Rekhi Centre of Excellence for the Science of Happiness, IIT Kharagpur, India*

[4]*School of Languages and Linguistics, Jadavpur University, India*

e-mail: uddaloksarkar@gmail.com

[Received: 01-12-2020; Revised: 31-12-2020; Accepted: 31-12-2020]

ABSTRACT

Music is often considered as the language of emotions. It has long been known to elicit emotions in human being and thus categorizing music based on the type of emotions they induce in human being is a very intriguing topic of research. When the task comes to classify emotions elicited by Indian Classical Music (ICM), it becomes much more challenging because of the inherent ambiguity associated with ICM. The fact that a single musical performance can evoke a variety of emotional response in the audience is implicit to the nature of ICM renditions. With the rapid advancements in the field of Deep Learning, this Music Emotion Recognition (MER) task is becoming more and more relevant and robust, hence can be applied to one of the most challenging test case i.e. classifying emotions elicited from ICM. In this paper we present a new dataset called ***JUMusEmoDB*** which presently has 400 audio clips (30 seconds each) where 200 clips correspond to happy emotions and the remaining 200 clips correspond to sad emotion. The initial annotations and emotional classification of the database has been done based on an emotional rating test (5-point Likert scale) performed by 100 participants. The clips have been taken from different conventional *'raga'* renditions played in sitar by an eminent maestro of ICM and digitized in 44.1 kHz frequency. The ragas, which are unique to ICM, are described as musical structures capable of inducing different moods or emotions. For supervised classification purposes, we have used 4 existing deep Convolutional Neural Network (CNN) based architectures (resnet18, mobilenet v2.0, squeezenet v1.0 and vgg16) on corresponding music spectrograms of the 2000 sub-clips (where every clip was segmented into 5 sub-clips of about 5 seconds each) which contain both time as well as frequency domain information. The initial results are quite inspiring, and we look forward to setting the baseline values for the dataset using this architecture. This type of CNN based classification algorithm using a rich corpus of Indian Classical Music is unique even in the global perspective and can be replicated in other modalities of music also. This dataset is still under development and we plan to include more data containing other emotional features as well. We plan to make the dataset publicly available soon.

© 2021 Acoustical Society of India

1. INTRODUCTION

Music imposes emotion. Considering the very basic, major scale always creates a happier ambience whereas minor scale creates a bit sad. This response to melody and rhythm is a biological instinct. A non-musician's ear can feel the pain when Garry Moore plays "Loner", even a baby responds to different scales and melody lines. While talking about emotions elicited by Indian Classical Music (ICM) it always becomes a matter of huge challenge for its ambiguous emotional response. Studies involving non-linear techniques have been conducted in the recent past to understand this complex behavior of music and its manifestation in the human brain[13-17].

In the recent years, Machine Learning has made significant advancements in a multitude of fields including computer vision, medical imaging, natural language processing and so on[18-37]. Such Machine Learning and Deep Learning approaches have been used to identify different emotions associated with music[9-12]. So, Music Emotion Recognition (MER) has always been an interesting task to perform for observing the correlation between the music and perceived emotion. Music Emotion Recognition task was first introduced by Barthet *et al.*[5]. Since then, several developments have been achieved on this discipline, and hence it has become very important to observe the role of ICM in elicitation of emotion.

To exploit the significance of emotion induction in ICM, we need a proper database to work with. Previously datasets like Computer Audition Lab 500-song (CAL500)[6], CAL500exp[7], datasets have been introduced, which is enriched with 500 western music clips. Here in this paper we have introduced a database ***JUMusEmoDB*** enriched with 400 music clips from genre of Indian Classical Music, of which 200 clips fall into the category of happy emotion while 200 falls into sad. Each clip is of 30 seconds length which is long enough for introducing an emotional imposition[8]. All the clips are parts of different 'raga' renditions improvised in Sitar by an eminent maestro of ICM. Each Raga in ICM evokes not a particular emotion (rasa), but a superposition of different emotional states such as joy, sadness, anger, disgust, fear *etc.* To decipher which particular emotion is conveyed in the chosen 30 sec segment of the raga, an emotion annotation was performed initially by 100 participants based on 5-point Likert scale.

For an emotion classification task different acoustic features of music are very important. Different acoustic feature consists of (a) Rhythmic Features: Tempo, Silence etc.; (b) Timbral Features: MFCC, Average Energy, Spectral Centroid, Spectral Tilt etc.; (c) Chroma Features. These acoustic features quantify the musicality of a clip which in effect contributes to MER. Study on this knowledge driven approaches results in an efficient model structure; but before this, a validated dataset is very necessary to work with. After the emotion annotation, a data driven approach has been used to validate our dataset.

We have taken an image processing-oriented approach to classify the dataset into emotion tags. We primarily extracted the spectrogram of a clip and fed the processed spectrogram image into existing deep Convolutional Neural Network (CNN) based architectures. These CNNs were then trained to classify emotions. For this study, we have made use of four different existing CNN architectures: VGG16, ResNet18, MobileNet v2.0, SqueezeNet v1.0 and have received some promising results. Furthermore, the dataset used in our study which we have named as JUMusEmoDB, is a novel dataset comprising of clips from Indian Classical Music genre. This dataset currently has musical clips from two emotions, namely happy and sad, and is still under development. We plan to include more data containing other emotional features eventually making the dataset publicly available shortly. This dataset can be used in future by the scientific community for emotion classification purposes, to investigate the impact of Indian Classical Music on human brain and finally to conduct cross-cultural studies combining both Western Classical and Indian Classical musical clips.

The paper has been organized as follows: Section 2 contains Data Acquisition, Section 3 comprises of the Methods used, the Experiments and Results of the study have been mentioned in Section 4 and finally Section 5 has the conclusion.

2. DATA ACQUISITION

JUMusEmoDB consists of 400 audio clips of 30 second each. The clips have been taken from different conventional raga renditions played in sitar by an eminent maestro of ICM and recorded with a sample frequency of 44.1 kHz. 200 clips correspond to happy emotions and the remaining 200 clips correspond to sad emotion. Initial annotations and emotional classification of the database has been done based on an emotional rating test (5-point Likert scale) performed by 100 participants.

3. METHODS

We have labeled our music database ***JUMusEmoDB*** into two main classes as stated earlier, *i.e.*, happy, and sad. In this paper we have followed the approach of a data driven method to classify the database into two distinct classes. Now a general question arises regarding why a data driven approach is being followed primarily rather than any novel knowledge-based quantification. To answer this, we have to take into account that a successful classification with high accuracy rate will validate the authenticity of ***JUMusEmoDB,*** which can then be used to develop new knowledge-based models to test on this database.

3.1 Data Preprocessing

The basic input to our CNN based framework should be an image data. Hence, to map our audio database into image paradigm we have made use of Spectrogram. But before obtaining spectrogram for these 30-second-long audio clips, we have sliced all the 30 second clips into individual 5 second clips to augment our dataset and performed STFT on this augmented dataset to obtain spectrograms.

3.2 Spectrogram

In our proposed framework, we have performed our classification task on the mel-spectrograms of derived music tracks. To extract the melspectrogram we have made use of short time Fourier transform with a window size of 2048 and hop size of 512 to obtain a spectrogram (Fig. 1). Thus we obtained melspectrogram as a dot product of obtained spectrogram with mel filterbanks.

$$spect(t,f) = |stft(t,f)|^2$$

$$melspectrogram(t,f') = spect(t,f) \; . \; melfilterbanks$$

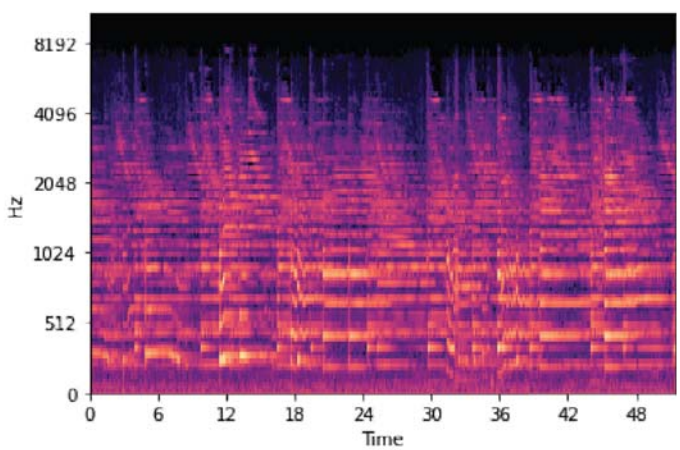


Fig. 1. Spectrogram of a HAPPY Music Clipping

3.3 CNN Models

We have extended our framework into four different established ConvNet models, *i.e.*, VGG[1], ResNet[2], SqueezeNet[3], MobileNet v2[4]. We thus have obtained four accuracy rates with an average of 99.117%. We have added an extra layer of 2 channels at each end of considered model because of bi-class output of our framework.

3.3.1 VGG Net is the oldest architecture among the used ConvNet models, proposed by Karen Simonyan and Andrew Zisserman of Oxford Robotics Institute in 2014[1]. This has sixteen weight layers; thirteen convolution layers divided into five groups, each group followed by pooling layers, and three fully connected (FC) layers at the end of whole network (Fig. 2). Convolution layers have a receptive field of 3×3 throughout the whole net, with stride 1. The Maxpooling layers consist of receptive fields of size 2×2 each and with a stride of 2. The network ends with three fully connected layers with first two layers of 4096 channels and last layer of 1000 channels due to 1000 classes of ILSVRC (ImageNet Large-Scale Visual Recognition Challenge).

Fig. 2. VGG-16 Network Architecture

3.3.2. ResNets are residual learning framework with substantially deeper network but with lower complexity. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun has proposed a residual network with a depth of up to 152 layers (Fig. 3-4) *i.e.*, 8x deeper than VGG-19 nets[2]. Increasing depth of a network can lead to a very serious problem of vanishing gradient which results in saturation of convergence with a very high training error and low accuracy problems. Kaiming He *et al.* has beautifully taken care of these facts and modified a very deep network to gain a high accuracy with low training error. They have implemented a *"short-cut connection"* of identity mapping. Their approach was to allow the network to fit the stacked layers to a residual mapping using residual block (Fig. 3) instead of fitting them directly to the underlying mapping. So, instead of feeding H(x) (desired underlying mapping) let the stacked network fit, $F(x) := H(x) - x$ and thus ultimately gives $H(x) := F(x) + x$.

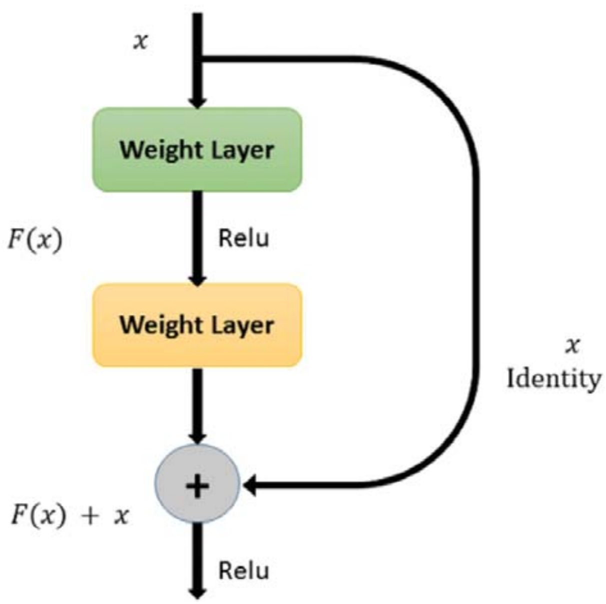


Fig. 3. Residual Block

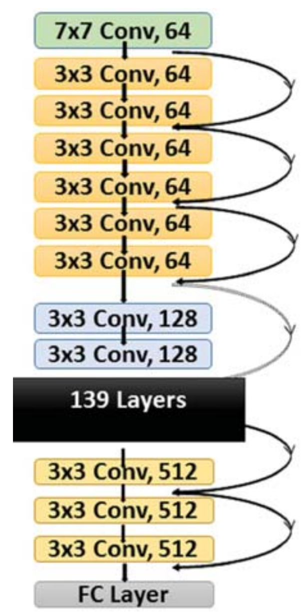


Fig. 4. ResNet 152-layer Architecture

3.3.3. SqueezeNet is a lighter modification of deep convolutional neural networks which has achieved an accuracy near to AlexNet (on ImageNet dataset) with 50 times fewer parameters (Fig. 6)[3]. The main concept behind this architecture is introduction of 'fire module". A fire module is a stacking of a squeeze layer with 1x1 convolution filters and an expand layer which has both 1x1 and 3x3 filters. Number of

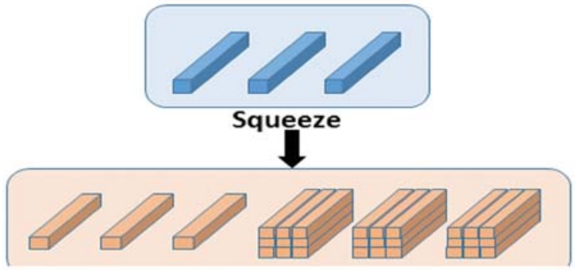


Fig. 5. Fire Module

kernels in squeeze layer should be less than number of kernels in expand layer to limit the number of input channels to 3x3 kernels. Fig. 5 shows an architectural view of fire module.

3.3.4. MobileNet, by Google, has introduced a new kind of lightweight architecture by replacing traditional convolution layer with "Depth-wise Separable Convolution" to reduce the model size and complexity. In MobileNetV2 two kinds of blocks are observed[4], stride 1 block (residual block), stride2 block (downsizing). Each block consists of three layers as shown in Fig. 7. First layer contains 1x1 kernel with RELU6. Second layer performs the depth-wise convolution. Third layer again contains 1x1 kernel without any nonlinear function.

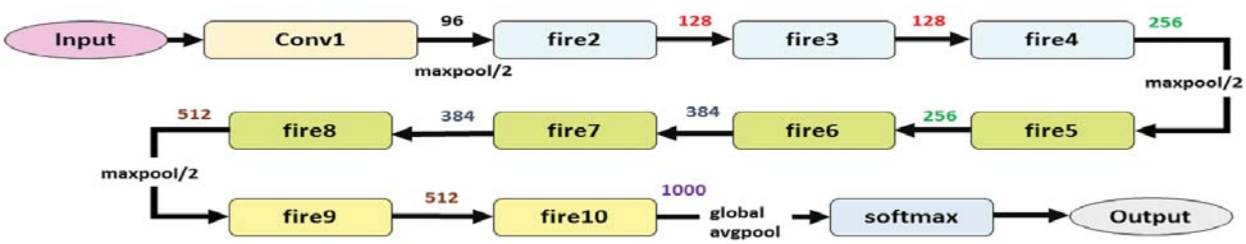


Fig. 6. SqueezeNet Architecture

3.4 Emotion Extraction from Network output

Output from the fully connected layers is provided to softmax to extract out the classes. In our framework, we are concerned about two distinct classes (*i.e.*, Happy Music clip, Sad Music clip). All the four architectures contain 1000 channels in last fully connected networks by default 1000 classes of ILSVRC (ImageNet Large-Scale Visual Recognition Challenge). So, we added another fully connected layer of 2 channels at the end to cater our purpose. After this last fully connected layer a softmax is performed to extract out the emotion classes.

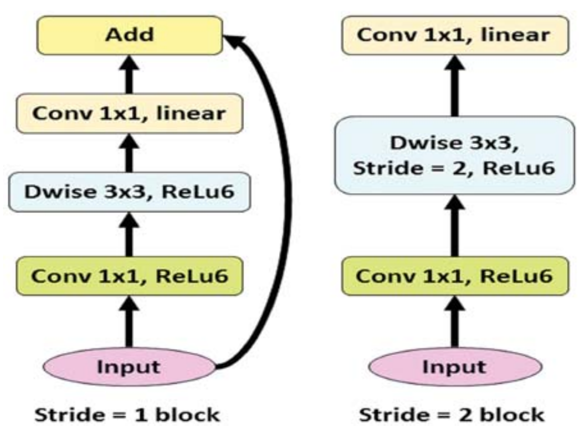


Fig. 7. MobileNet Architecture

4. EXPERIMENTS AND RESULTS

As a classification task, some criteria are being used to quantify the classification performance. MSE (mean squared error), MAE (mean absolute error) has been used a lot as loss functions for image purposes. In our classification task we have made use of cross entropy loss. Entropy is a measure of uncertainty, and it is measured as,

$$H(X)=\begin{cases} -\int p(x)\log(p(x)), & x \text{ is continuos random variable} \\ -\sum -\int p(x)\log(p(x)), & x \text{ is discrete random variable} \end{cases}$$

Cross entropy loss (by using the idea of Entropy) measures the resemblance of actual output against predicted output. Cross entropy increases along with the divergence of prediction from actual output. Hence a 0 loss represents perfect model. Cross entropy loss or Log loss is being calculated using the following equation:

$$L=-\sum_i y_i log(\hat{y}_i)w$$

where y_i is the calculated or predicted output and $(\hat{y}_i)$ is the actual output.

Previously we have introduced our used framework and model improvisation. In this section we are illustrating the acquired results. For our work, we used a train/validation split ratio as 85/15. In the training phase we receive a model cost (calculated from loss function) which indicates the model performance. The loss curve for training phase for each model is shown in the Fig. 8. With time (iterations) convergence is achieved for each of the models used and the performance of SquuezeNetV1 in the training phase also outperforms the other three models used.

The convergence plot helps to tune the parameters of CNN models. After adjusting the parameters, we obtain the best results for each CNN Models. Table 1 shows the validation accuracy of the aforementioned models.

Table 1. Accuracies of different CNN models on the validation dataset

Model	Accuracy
VGG16	99.007%
ResNet18	97.682%
SqueezeNetV1	99.669%
MobileNetV2	98.675%

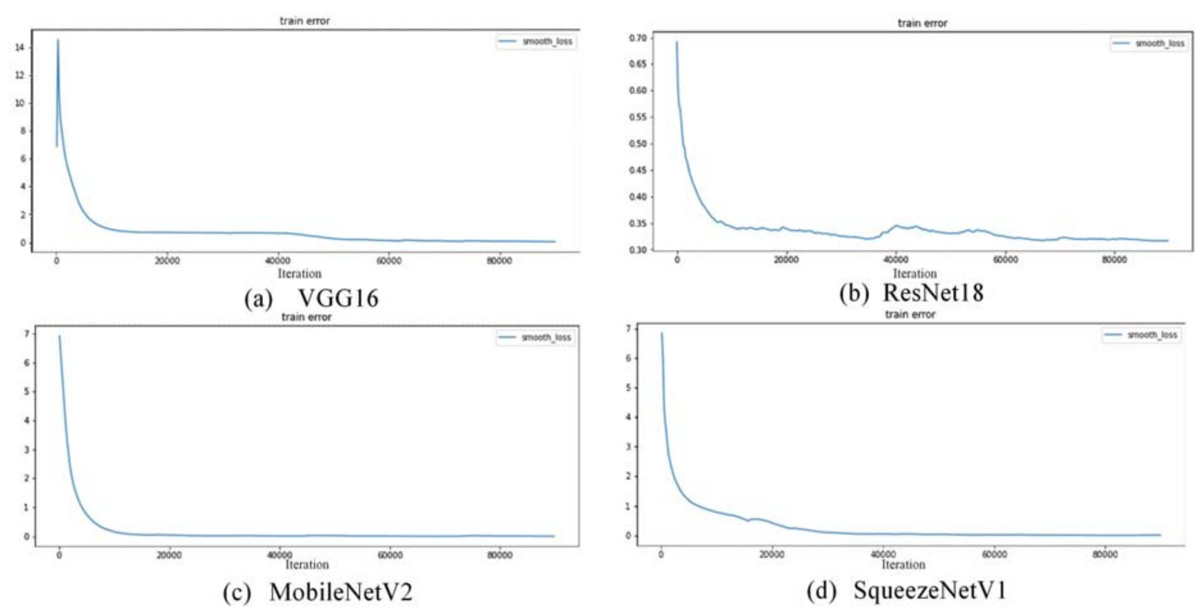


Fig. 8. Training Loss curves

From Table 1, we can see that SqueezeNetV1 seems to fit best with the acquired dataset. The dataset size (as of now) is not that huge which makes it suitable for light weight models (with fewer parameters) like SqueezeNetV1 which gives the best validation accuracy.

4. CONCLUSION

In this work, we proposed a novel dataset called JUMusEmoDB which presently has 400 audio clips (30 seconds each) where 200 clips correspond to happy emotions and the remaining 200 clips correspond to sad emotion. The initial annotations and emotional classification of the database has been done based on an emotional rating test (5-point Likert scale) performed by 100 participants. We also demonstrated the performances of four deep CNN based architectures namely resnet18, mobilenet v2.0, squeezenet v1.0 and vgg16. Validation accuracy values showed that SqueezeNetV1 performed the best out of the four models. Even though the advantage of employing CNN based architectures for tackling the problem of music emotion recognition include better overall accuracy because of better extraction of useful features from the data compared to other traditional methods, further studies need to be conducted to understand the source of emotions for a given music. Another limitation of this work is the lack of data in our dataset, but this is a pilot study it is still under development. We plan to incorporate more data containing other emotional features as well and eventually make the dataset publicly available shortly.

5. ACKNOWLEDGEMENT

Archi Banerjee acknowledges the Department of Science and Technology (DST), Govt. of India for providing the DST CSRI Post Doctoral Fellowship (DST/CSRI/PDF-34/2018) to pursue this research work. Shankha Sanyal acknowledges DST CSRI, Govt of India for providing the funds related to this Major Research Project (DST/CSRI/2018/78 (G)) and the Acoustical Society of America (ASA) for providing the International Students Grant.

6. REFERENCES

[1] Simonyan, Karen and Andrew Zisserman, 2014. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556.*

[2] He Kaiming *et al.* 2016. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*.

[3] Iandola, Forrest N. *et al.* 2016. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size." *arXiv preprint arXiv:1602.07360*.

[4] Sandler Mark *et al.* 2018. "Mobilenetv2: Inverted residuals and linear bottlenecks." *Proceedings of the IEEE conference on computer vision and pattern recognition*.

[5] Barthet Mathieu, György Fazekas and Mark Sandler, 2012. "Music emotion recognition: From content-to context-based models." *International Symposium on Computer Music Modeling and Retrieval.* Springer, Berlin, Heidelberg.

[6] Turnbull Douglas, *et al.* 2007. "Towards musical query-by-semantic-description using the cal500 data set." *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval.*

[7] Wang Shuo-Yang *et al.* 2014. "Towards time-varying music auto-tagging based on CAL500 expansion." *2014 IEEE International Conference on Multimedia and Expo (ICME). IEEE.*

[8] Turnbull Douglas *et al.* 2007. "Towards musical query-by-semantic-description using the cal500 data set." *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval.*

[9] Juthi J.H., Gomes A., Bhuiyan T. and Mahmud I. 2020. Music Emotion Recognition with the Extraction of Audio Features Using Machine Learning Approaches. In: Singh P., Panigrahi B., Suryadevara N., Sharma S., Singh A. (eds) Proceedings of ICETIT 2019. *Lecture Notes in Electrical Engineering,* vol **605**. Springer, Cham. https://doi.org/10.1007/978-3-030-30577-2_27

[10] Liu Xin *et al.*, 2017. "CNN based music emotion classification." arXiv preprint arXiv:1704.05665.

[11] Liu Tong *et al.*, 2018. "Audio-based deep music emotion recognition." AIP Conference Proceedings. *AIP Publishing LLC*. **1967**(1).

[12] Yang Yi-Hsuan and Homer H. Chen, 2012. "Machine recognition of music emotion: A review." *ACM Transactions on Intelligent Systems and Technology (TIST)* **3.3**, 1-30.

[13] Sanyal Shankha, *et al.*, 2019. "Music of brain and music on brain: a novel EEG sonification approach." *Cognitive neurodynamics* **13.1**, 13-31.

[14] Sengupta Sourya *et al.*, 2017. "Emotion specification from musical stimuli: An EEG study with AFA and DFA." *4th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE.*

[15] Nag Sayan *et al.*, 2017. "Can musical emotion be quantified with neural jitter or shimmer? A novel EEG based study with Hindustani classical music." *4th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE.*

[16] Sarkar Uddalok, *et al.,* 2020. "A Simultaneous EEG and EMG Study to Quantify Emotions from Hindustani Classical Music." *Recent Developments in Acoustics. Springer, Singapore.* pp. 285-299.

[17] Sayan Nag, Uddalok Sarkar, Shankha Sanyal, Archi Banerjee, Souparno Roy. Samir Karmakar, Ranjan Sengupta and Dipak Ghosh, 2019. A Fractal Approach to Characterize Emotions in Audio and Visual Domain: a Study on Cross-Modal Interaction. *Journal of Image Processing & Pattern Recognition Progress.* **6**(3), 1-7.

[18] Krizhevsky A., Sutskever I. and Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. *In Advances in neural information processing systems,* pp. 1097-1105.

[19] LeCun Y., 2015. LeNet-5, convolutional neural networks. URL: http://yann. lecun. com/exdb/lenet, 20.

[20] LeCun Yann, Yoshua Bengio and Geoffrey Hinton, 2015. "Deep learning." nature 521.7553 pp. 436-444.

[21] Goodfellow Ian *et al.*, 2014. "Generative adversarial nets." *Advances in neural information processing systems.*

[22] Goodfellow Ian *et al.*, 2016. *Deep learning*. Cambridge: MIT press. **1**(2).

[23] Xu Kelvin *et al.*, 2015. "Show, attend and tell: Neural image caption generation with visual attention." *International conference on machine learning.*

[24] Kingma, Diederik P. and Max Welling, 2013. "Auto-encoding variational bayes." arXiv preprint arXiv:1312.6114.

[25] Baldi P., 2012. Autoencoders, unsupervised learning, and deep architectures. *In Proceedings of ICML Workshop on Unsupervised and Transfer Learning,* pp. 37-49.

[26] Ronneberger Olaf, Philipp Fischer and Thomas Brox, 2015. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham.

[27] Hammernik Kerstin *et al.*, 2018. "Learning a variational network for reconstruction of accelerated MRI data." *Magnetic resonance in medicine* **79.6**, 3055-3071.

[28] Bhattacharyya Mayukh and Sayan Nag, 2019. "Hybrid Style Siamese Network: Incorporating style loss in complimentary apparels retrieval." arXiv preprint arXiv:1912.05014.

[29] Pu Yunchen *et al.*, 2016. "Variational autoencoder for deep learning of images, labels and captions." *Advances in neural information processing systems*.

[30] Nag Sayan, 2020. "Lookahead Optimizer Improves the Performance of Convolutional Autoencoders for Reconstruction of Natural Images." OSF Preprints. doi:10.31219/osf.io/95wge.

[31] Alex Krizhevsky and Geoffrey Hinton, 2009. Learning multiple layers of features from tiny images. *Technical report, University of Toronto.*

[32] Vincent Pascal *et al.*, 2010. "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion." *Journal of machine learning research* 11.12.

[33] Pu Yunchen *et al.*, 2016. "Variational autoencoder for deep learning of images, labels and captions." *Advances in neural information processing systems*.

[34] Vincent Pascal *et al.*, 2008. "Extracting and composing robust features with denoising autoencoders." *Proceedings of the 25th international conference on Machine learning*.

[35] Hochreiter Sepp and Jürgen Schmidhuber, 1997. "Long short-term memory." *Neural computation* **9.8**, 1735-1780.

[36] Vaswani Ashish *et al.*, 2017. "Attention is all you need." *Advances in neural information processing systems,* **30**, 5998-6008.

[37] Brown Tom B. *et al.*, 2020. "Language models are few-shot learners." arXiv preprint arXiv:2005.14165.

The Journal of Acoustical Society of India : Vol. 48, No. 1-2, 2021 (pp. 20-34)

Automatic stroke classification of tabla accompaniment in Hindustani vocal concert audio

Rohit M.A. and Preeti Rao
Department of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai, India
e-mail: {rohitma, prao}@ee.iitb.ac.in

[Received: 15-12-2020; Revised: 31-12-2020; Accepted: 31-12-2020]

ABSTRACT

The tabla is a unique percussion instrument due to the combined harmonic and percussive nature of its timbre, and the contrasting harmonic frequency ranges of its two drums. This allows a tabla player to uniquely emphasize parts of the rhythmic cycle *(theka)* in order to mark the salient positions. An analysis of the loudness dynamics and timing deviations at various cycle positions is an important part of musicological studies on the expressivity in tabla accompaniment. To achieve this at a corpus-level, and not restrict it to the few recordings that manual annotation can afford, it is helpful to have access to an automatic tabla transcription system. Although a few systems have been built by training models on labeled tabla strokes, the achieved accuracy does not necessarily carry over to unseen instruments. In this article, we report our work towards building an instrument-independent stroke classification system for accompaniment tabla based on the more easily available tabla solo audio tracks. We present acoustic features that capture the distinctive characteristics of tabla strokes and build an automatic system to predict the label as one of a reduced, but musicologically motivated, target set of four stroke categories. To address the lack of sufficient labeled training data, we turn to common data augmentation methods and find the use of pitch-shifting based augmentation to be most promising. We then analyse the important features and highlight the problem of their instrument-dependence while motivating the use of more task-specific data augmentation strategies to improve the diversity of training data.

1. INTRODUCTION

Hindustani classical music is a genre of art music in India encompassing several sub-genres and forms of presentation. *Khyal*, one of the more modern forms, with roots in Dhrupad, consists of vocals as the lead, the harmonium (or sometimes the saarangi) for melodic accompaniment, the tabla for percussion accompaniment, and the tanpura in the background for the harmonic drone[1]. The performance opens with a short unmetered raga improvisation *(alap)* by the vocalist, performed to a mild accompaniment by the harmonium, and is followed by the metered piece - the bandish, where the tabla also joins in. For the most part, the melodic accompaniment shadows the vocals while the tabla plays a fairly fixed cyclic pattern of beats called the 'theka' to provide the rhythm.

© 2021 Acoustical Society of India

The tabla is a pitched percussive membranophone, consisting of a pair of drums called the *bayan* and *dayan*, or *dagga* and *tabla*, respectively. A unique property of these drums that sets them apart from common percussion instruments is their ability to produce ringing, pitched sounds. This is facilitated by the additional metallic layer at the center of each drum's membrane. When struck freely, the bayan, bigger in size, produces 1 bass frequency sounds (F0[1] $\in$ 80-100 Hz), while the dayan produces treble frequency sounds (F0 $\in$ 200-400 Hz). We therefore refer to these drums as bass and treble in the remainder of this article. The treble drum, tuned to the tonic note of the singing in a performance, is more harmonic than the bass drum which evokes only a faint sense of pitch.

Due to the predominantly improvisatory nature of Hindustani music, and a lack of precise performance oriented theory, a greater emphasis has come to be placed on the empirical analysis of performed concerts. While traditional musicological studies would often be restricted to a few chosen concerts, the advancement of computational methods has made it possible to analyse large corpora. The focus of some of these studies has been to quantify aspects of rhythm and tempo in khyal and instrumental music concerts from the perspective of the percussion accompaniment, such as - stroke loudness dynamics and beat-level timing deviations in the tabla playing[2], typical tempo ranges in khyal performance and the variations in theka based on the chosen tempo[3], and the possible interaction of this with the accompanied lead vocals or melodic instrument. There is thus a large scope to help further such research by making use of automatic tools for accurate tabla transcription to aid large-scale analyses.

Supervised methods in machine learning achieve great success when provided suficient labeled data that is matched to the target domain. In the present context, our target data, *viz.* tabla accompaniment from vocal concerts is available as such only in specially created multi-track recordings where each instrument is recorded with a separate microphone with suficient physical separation between the artists. Further, although musical source separation is a rapidly developing area, there is no model currently available for the Hindustani vocal concert context. So we develop our models on the more amply available tabla solo, hoping to extend the trained models to accompaniment tabla obtained from multi-track recordings or possibly future source separation techniques. In the next section, we review the previous work in tabla stroke classification. We then present our problem and dataset, followed by a discussion of our method and a critical evaluation of its performance on our test data.

2. BACKGROUND

The approach to transcribing tabla most commonly found in literature is one where segments of individual strokes are first represented using low-level audio features and then classified into one of a set of 10 - 15 bols [4–9]. The stroke segments are obtained either by extracting them from a sequence of tabla strokes using a separate segmentation method[4, 5, 8], or by building a dataset of individual strokes recorded in isolation[6, 9]. The various systems reported so far can be compared via the following aspects - the set of features used to characterize strokes, the models used for classification, the nature and diversity of the dataset, and the evaluation methods.

There is not a significant difference in the choice of features, with a majority of the previous methods employing a similar set of somewhat generic timbral descriptors such as Mel-Frequency Cepstral Coeficients (MFCC), spectral distribution features like the spectral centroid, skewness, and kurtosis, Linear Prediction Coeficients (LPC), and temporal features like the zero crossing rate, temporal centroid, and attack time. A dimensionality reduction step has been additionally used to select the most important subset of these features[5, 6]. However, while it would appear, from the fairly high classification accuracies reported, that these common low-level timbre descriptors are well-suited to the task, it is important to note that these descriptors tend to also depend on the physical characteristics of the tabla and differences in stroke articulation[10]. Therefore, along with analysing the quality of features based on the resulting separability of strokes[7], it is also important to consider the instrument independence of these features.

[1] Fundamental frequency

For the task of classification, a number of common models have been used, ranging from simple decision trees, k-nearest neighbours, and single layer feed-forward neural networks, to the slightly more complex random forests, probabilistic neural networks, and hidden markov models (HMM). The use of an HMM, which models temporal context information, seems to have helped in cases where ambiguous stroke labels were present in the dataset[4, 11]. Such ambiguities arise commonly in recordings of tabla compositions, where similar sounding strokes sometimes have different labels and *vice-versa*. In other cases, where the labels were based more on the stroke acoustics, or where the dataset consisted of isolated strokes, high classification accuracies have been achieved even without modelling the context.

The task of transcribing strokes drawn from continuous playing warrants a consideration of a few additional points over isolated stroke recordings. One, regarding the use of language models, is whether temporal context information learnt from solo compositions would be appropriate for tabla accompaniment transcription. While training language models on the basic thekas is a possibility, this would not capture all of the expressivity observed in concerts. An additional challenge is the overlap of harmonics from a previous stroke with subsequent ones[12]. For instance, a damped stroke like 'ke' (played on the bass drum) could be mistaken for a stroke like 'Na' with a sustained sound (produced on the treble drum) occurring just before it, because of the harmonics from 'Na' that continue to sound during 'ke'. Hence, a system trained on isolated strokes is likely to be of limited value in analysing tabla playing in concert audios.

The evaluation method is important not only in assessing the classification accuracy of a proposed method but more importantly, its generalisation to unseen instruments. Due to the drastic timbre differences between tabla sets arising from diverse physical properties and tuning, building a robust classifier that works equally well on a variety of instruments has been found to be a difficult task[5]. This is evident from the high cross-validation scores when models are evaluated on strokes from a tabla that they have already seen during training, and the significantly poorer scores on an unseen tabla. Therefore, it is necessary to not only build a diverse dataset, but also to evaluate the instrument-independence of the proposed methods.

Regarding the specific task of transcribing tabla accompaniment, there has been a recent attempt at identifying strokes in the theka of a few talas[8]. The dataset used is a diverse mix of several tabla sets, players, talas, and tempi. However, the recordings are of the thekas in their prescribed format, and limited in terms of fillers and extempore variations that we would find in a real concert. Further, while the results reported do reflect the challenges in transcribing stroke sequences as opposed to strokes recorded in isolation, there is no instrument independent evaluation reported.

3. PROBLEM FORMULATION

The main goal of the current work is to aid the analysis of tabla accompaniment in vocal concerts by providing musicologically meaningful stroke labels. While the outputs of a tabla transcription system can encompass all the distinct tabla bols, an important level in the taxonomy of bols is the classification into resonant and damped categories[13] (Figure 1). This is based on the acoustic characteristics of the sound produced, which in turn depends on the manner in which the drums are struck - resonant sounds, produced using impulsive strikes, are harmonic and sustained, whereas damped strokes are transient and are produced by preventing the membrane from continuing to vibrate after the strike. More

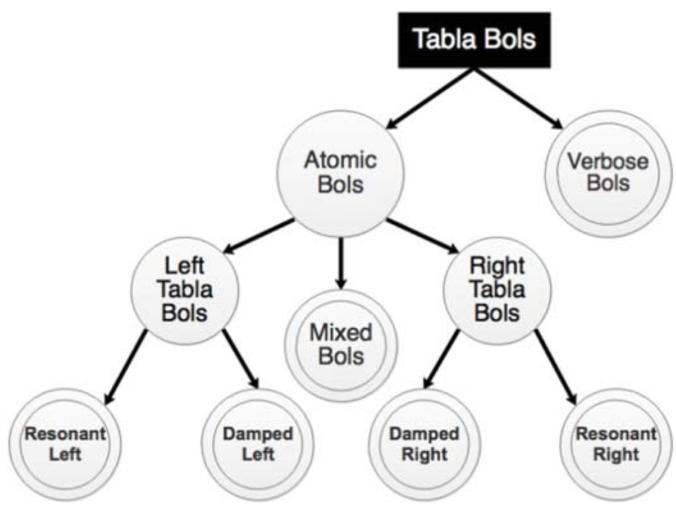


Fig. 1. Classification of tabla bols based on the drum that is struck and the nature of the sound produced[13]

importantly, these categories also have musically meaningful roles in the context of accompaniment, where they help mark different sub-divisions and accent positions in the theka[2]. For example, in the theka of *tintal*, a 16-beat cycle, there are four sub-divisions, each of 4 beats, which are primarily distinguished based on the presence or absence of the bass resonant sound. In contrast, the different sub-divisions of the theka of *ektal*, a 12-beat cycle, are distinguished based on whether they contain damped or resonant strokes.

We therefore look at transcribing tabla into the following four categories **- Damped, Resonant Treble, Resonant Bass** and **Resonant Both,** based on which drum is struck and whether the sound is resonant or damped. The damped category refers to strokes that do not have a sustained decay and includes all such strokes produced either by striking one of the drums or both simultaneously (each producing a damped stroke). The resonant treble and resonant bass categories correspond to strokes that produce sustained sounds on the corresponding drums (while the other drum either remains silent or produces a damped sound). Resonant both refers to strokes that are played simultaneously on both the drums, each of which results in a sustained sound. The acoustic characteristics of each category are described in Table 1, and Figure 2 shows the spectrograms for an example stroke of each category. Although a transcription into the exhaustive list of bols could be followed by then mapping the labels to any reduced set of categories, a more reasonable and also useful formulation of the problem, given the scarcity of data, is to classify each stroke directly into the required categories.

Table 1. The four target tabla stroke categories of the transcription and their acoustic characteristics.

Stroke Category	Acoustic Characteristics
Damped (D) *(Ti, Ta, Tak, Ke, Tra, Kda)*	No sustained harmonics, only a burst of energy at the onset
Resonant Treble (RT) *(Na, Tin, Tun)*	One or more strong harmonics in the 200-2000 Hz range persist after the onset
Resonant Bass (RB) *(Ghe, Dhe, Dhi)*	Usually a single strong harmonic close to 100 Hz persists after the onset
Resonant Both (B) *(Dha, Dhin)*	Both the above characteristics of resonant treble and bass

4. DATASET DESCRIPTION

The test data for our experiments was carefully chosen to closely resemble a realistic accompaniment scenario. Since most public concert audios are not available in a multi-track format with instruments recorded in perfect isolation, it was necessary for us to build such a dataset. For this, we first obtained solo singing audios from expert singers, and then got expert tabla players to play the corresponding tabla accompaniment while listening to the pre-recorded vocals over headphones. There are a total of 10 such aligned vocal-tabla audio pairs, spanning a net duration of nearly 20 minutes, and yielding about 4500 strokes. 6 of the compositions are in *madhya lay tintal* (130 - 160 BPM), and the rest are in one each of *jhaptal* and *bhajani theka* in *madhya lay*, and *tintal* and *dadra* in *drut lay* (≈ 250 BPM). The accompaniment was played by 2 different musicians for half of the tracks each, with one of the players using 2 different tabla sets based on the tonic of the singing (there are thus a total of 3 distinct tabla sets each of different tuning). All the tabla audios were recorded at a sampling rate of 44.1 kHz in quiet environments using professional microphones.

On the other hand, given the expensive nature of collecting such data, we obtained the data for training our classification model from three available sources of tabla solo playing, each recorded in the absence of a lehra (melodic accompaniment), as described in Table 2. The PP and AS subsets consist of about 20 recordings each of short tabla solo compositions (5 of which are common), with each track between 10 and 30 seconds long. The AHK subset has seven short excerpts from a solo concert recording, each between 30 and 60 seconds long. The recording, taken from YouTube[2], is of poorer quality and at a lower sampling

[2] https://www.youtube.com/watch?v=mEFr1Tp801M

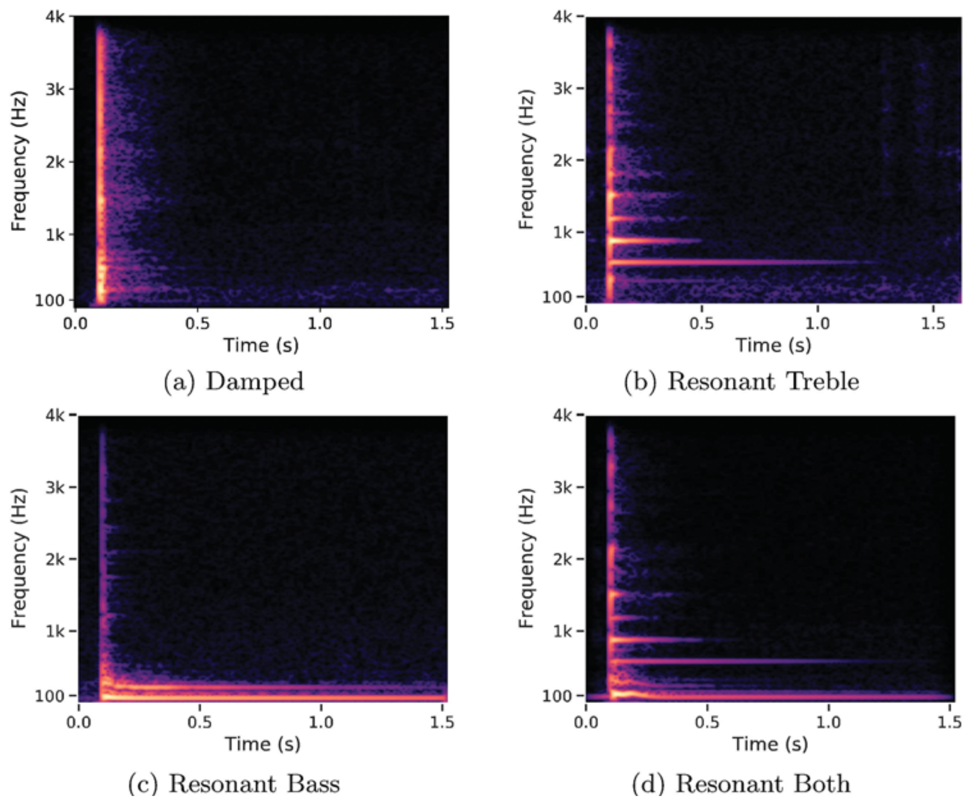


Fig. 2. Magnitude spectrograms of a stroke from each of the four target stroke categories (spectrum shown up to 4 kHz for better visualisation of the strong harmonics whose frequency support distinguishes the resonant bass and treble strokes)

Table 2. Description of the training *(PP, AS, AHK)* and testing datasets.

Name	Source	Nature of playing	Duration	# strokes
PP	Rohit and Rao (2018)[14]	Solo compositions	7 mins.	2848
AS	Rohit and Rao (2019)[15]	Solo compositions	5 mins.	1432
AHK	Tabla solo concert audio from YouTube	Solo compositions interspersed with theka	6 mins.	2400
Test	Recordings made for this study	Isolated accompaniment to pre-recorded vocals	20 mins.	4470

rate of 16 kHz. The excerpts were chosen by excluding portions of extremely fast playing that are difficult to annotate. Each of these subsets corresponds to a single unique tabla set and player.

The *PP* and *AS* subsets are available along with time-aligned bol annotations. These bols were mapped according to their acoustic characteristics to the four stroke categories, for the purpose of this study. As for the *AHK* and test sets, the annotation was performed manually with the help of Audacity[3] by the first author (a tabla student) and an intern in the same lab. To aid the manual annotation process, stroke onsets were first determined automatically using an automatic onset detection algorithm (Section 5.1). In order

[3] Audacity ® software is copyright © 1999-2020 Audacity Team. The name Audacity ® is a registered trademark of Dominic Mazzoni.

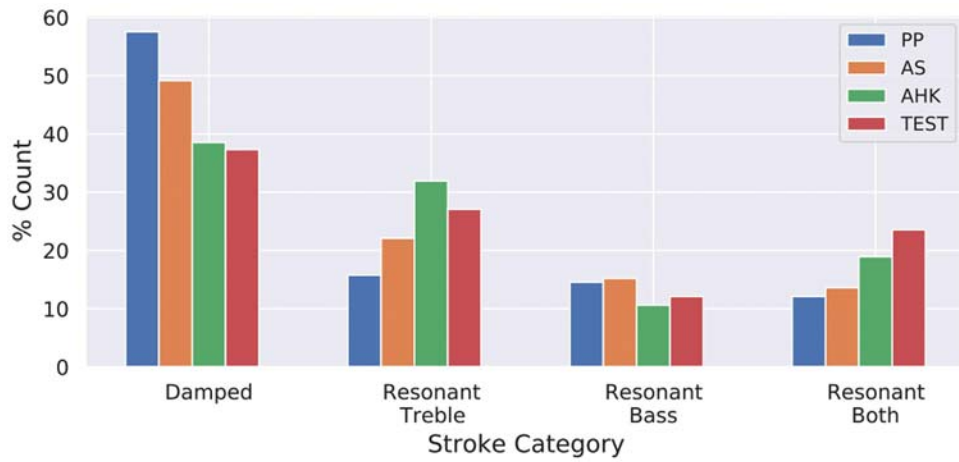


Fig. 3. Distribution (% count) of strokes across categories in the training and testing sets.

to make up for missed soft onsets, a fairly low threshold was used while picking the peaks in the onset detection signal. This resulted in some false alarms, but ensured that soft onsets were also detected, thus requiring the annotator to not add any onsets, but only remove the false alarms and appropriately label the correct onsets.

The distribution of strokes across the four categories in each of these subsets is shown in Figure 3. Comparing these with the test set, we see a few similarities as well as differences. In all the subsets (train and test), the damped category is the most populated and resonant bass is the least populated. The high count of damped strokes can be attributed, in the case of solo playing, to the nature of solo compositions - they contain more damped strokes to allow playing at high speeds, and in the case of accompaniment, to fillers and expressive embellishments. Further, in the test set, the resonant both category has a higher count than any of the training subsets, possibly because of the dominant nature of such strokes in accompaniment playing (for instance, the theka for tintal has 12 out of its 16 strokes of this kind).

5. METHODS

In the present study, we follow the same ‘segment and classify’ approach, where stroke onsets are first determined and the labels are then assigned to the segmented strokes. Each segment is the duration from the onset of a stroke to the onset of the subsequent stroke. We extract features on these segments and provide the feature vectors to a random forest classifier model to predict the stroke category. All the tabla audios (training and testing) are first down-sampled to 16 kHz, since the *AHK* set is not available at 44.1 kHz.

5.1 Onset Detection

Given the well-studied nature of onset detection, especially for percussive sounds, we make use of a common onset detection method called the High Frequency Content, and experiment with the post-processing hyperparameters to arrive at the best settings for our data. This method involves calculating the frequency weighted sum of the magnitude spectrum to indicate the onset strength in each frame[16]. The implementation is as provided in the software library *Essentia*[17], with the window and hop size values set to 25 ms and 5 ms, respectively. The post-processing step involves smoothing the onset detection signal and then picking the peaks above a certain threshold[18]. The size of the moving average filter *(delay)* used for smoothing and the threshold value for peak-picking *(alpha)* are hyperparameters that we experiment with. We first iteratively evaluate each of the methods on the training set using a grid of values for *alpha* and *delay*, and pick the pair of values that results in the highest f-score. Then, we keep these values fixed and evaluate the onset detector on the test set.

5.2 Feature Extraction

The features we use in this work are modeled on those used in previous tabla transcription methods (as reviewed in Section 2), along with a few important modifications and additions. The total set of features, 49 in all, is given in Table 3. Some of the features are calculated separately on band-passed versions of the signal. Given that our four-way target stroke classes are based closely on the differences in the resonant characteristics of each drum, we decide on the following ranges for the bands: 50 - 200 Hz to capture the base and 200 - 2000 Hz to capture the treble harmonics. These ranges are kept broad in order to capture a wide F0 range of each drum, similar to the bands used in[4]. They could, however, be tuned based on the tabla as well, by first identifying the range of harmonic frequencies of each drum, as done previously (for the treble drum alone) in[13]. This could be achieved by analysing a few resonant strokes of each drum, which, in a testing scenario, could be provided manually.

Table 3. The set of features used for stroke classification.

Category	Feature	Count	
Spectral	Spectral Centroid, Skewness, Kurtosis	2 (mean, stdev.)	
	MFCC	13 (mean)	
	Flux (onset strength)	1 (max.)	x 2 (bass, treble)
	Energy	3 (sum, mean, stdev.)	x 2 (bass, treble)
Temporal	Log Attack Time	1	
	Temporal Centroid	1	
	Zero Crossing Rate	2 (mean, stdev.)	
	Early Decay Rate & Intercept	2	x 2 (bass, treble)
	Late Decay Rate & Intercept	2	x 2 (bass, treble)
	R^2	1	x 2 (bass, treble)
	Spline Knot Location	1	x 2 (bass, treble)
Delta	Sum and Mean Energy	2	x 2 (bass, treble)
	Late Decay Rate	1	x 2 (bass, treble)
	Total	**49**	

All the spectral features, and the temporal zero-crossing rate feature, are calculated frame-wise on overlapping short-time frames of size 25 ms, with a hop of 5 ms, over the entire stroke segment. The shape-related spectral features (spectral centroid, skewness, and kurtosis) are summarised using the mean and standard deviation, MFCC using the mean, flux using the maximum, and energy using the sum, mean, and standard deviation. The other two temporal features - log attack time and temporal centroid are calculated from the amplitude envelope of the entire time-domain audio signal of the stroke segment. The implementation provided by Essentia is used for each of these features.

For the remaining temporal features, the short-time energy envelope serves as the base. A logarithmic transformation is first applied to make the decay portion (the part beyond the peak) of the envelope more linear. By observing a few plots of band-wise log-scaled short-time energy envelopes (Figure 4), we see that the decay portion usually consists of two parts (early and late), and the differences between the stroke categories seem to be based on the rates of these decays. For instance, the decay patterns in both the bands are quite similar for the damped stroke, with the late portion mostly corresponding to noise, whereas in the resonant both stroke, the bands contain smooth decays at different rates. Therefore, to capture the distinctive nature of the decay, a linear spline model is fit to the decay portion. The 'knot' location for the

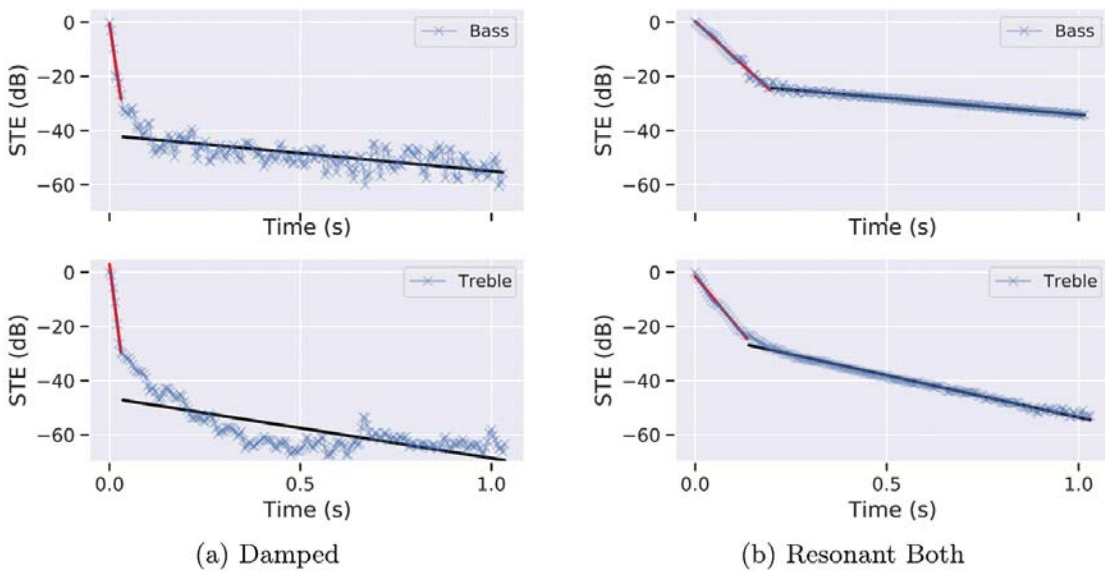


Fig. 4. Band-wise log-transformed short-time energy envelopes (marked by 'x') of a stroke from (a) Damped and (b) Resonant Both categories, along with the best spline fit (solid lines).

fit, which is the point of inflexion, is automatically estimated as the point that yields the combined best fit for the two linear segments, in the sense of the harmonic mean of the R^2 values of each fit. The slopes of both the line segments, their y-intercepts, the harmonic mean of the R^2 values, and the knot location are all used as features. The terms 'early' and 'late' are used to refer to the two line segments in terms of their relative positions. The use of decay-related parameters as features is reported in[13]. However, a key difference here is the use of a linear spline fit to the log-scaled envelope instead of a single exponential fit to the un-scaled envelope as done previously.

Finally, to help determine whether the measured characteristics of a stroke are not corrupted by possible overlap from a previous stroke, we incorporate delta features. These are calculated as the difference between the values in a given and its previous stroke for the following six features - bass and treble late decay rates, and the sum and mean of base and treble short-time energy values.

5.3 Stroke Classification

For stroke classification, we use a random forest classifier, with the implementation provided in the scikit-learn library[19]. A random forest classifier is an ensemble of decision trees in which the output is determined by a majority vote across all the trees. The main motivation behind its design is to prevent the overfitting that is commonly observed in the case of a single decision tree. This model is also useful because it lends itself well to an analysis of the important features in the input data (given by the 'feature importances' attribute of a trained model). The input to the model is a feature vector comprising all the features listed in Table 3, extracted on a single stroke segment, and the target is one of the four stroke categories. For our initial experiments, we set the number of trees in the forest to 200, the 'max features' parameter to 'sqrt', and the 'bootstrap' parameter to 'True' (Section 5.3.1).

5.3.1 Feature Selection and Hyperparameter Tuning : The subset of features that are most relevant and non-redundant can be obtained via a recursive feature elimination (RFE) strategy. This method uses a classifier like the random forest to first estimate the relative feature importances, and then removes the least important ones recursively until a desired number of features is obtained. In order to obtain the optimum number of features, it is performed in a cross-validation setting on the training set, where a second classifier model is trained and evaluated on separate folds to obtain the average cross-validation accuracy at a given number of selected features. By repeating this for different values of the number of selected features and recording the average classification score in each case, a suitable point on this curve can be chosen while trading off performance for complexity.

An important step in extracting the best performance out of a classifier is the optimization of its hyperparameters. In the case of the random forest, these determine various aspects of the model complexity and help control the trade-off between bias and variance - a more complex model increases the variance while a simpler model increases the bias. In the present work, we look at optimizing the following hyperparameters - number of trees, maximum allowed depth of each tree, size of the subset of features given to each tree (also called 'max features'), and the use of bootstrapping in sampling subsets of the training data. The optimization is carried out by sampling 40 random combinations of these hyperparameters from a grid of values and evaluating each combination in terms of the average cross-validation score on the training set. The use of cross-validation here is intended to not bias the model towards our test set of accompaniment audios in the process of tuning the model hyperparameters.

5.3.2. Data Augmentation and Balancing : A recurring problem that has been noted by authors in some of the previous work on tabla transcription is the scarcity and limited diversity of available datasets. The process of building large, accurately annotated datasets is, in general, extremely cumbersome, but especially so in the case of instruments like tabla, which cannot be synthesized realistically using existing software or electronic instruments (most software only has recorded samples of short loops). This motivates the use of effective data augmentation strategies where available labeled data is transformed in different ways that do not change the validity of the labels, thus offsetting to an extent the need to collect more data.

For our present work, we investigate the following three general data balancing and augmentation methods - repeating the data samples (repeated oversampling), interpolating in the feature space, and pitch-shifting the audios. For pitch-shifting, we use the function provided in the software library *librosa*[20], and only consider the three resonant stroke categories (since damped strokes do not contain a pitched component). Each stroke's audio signal is pitch-shifted by each of 4 semitone levels: {-0.5, -0.25, 0.25, 0.5}. This rather small range of shifts compares well with the allowed range for a high-pitched tabla of a small-diameter, which is what the tabla sets in our training set resemble[21].

Given that our dataset is also highly imbalanced in terms of the distribution of examples across the four categories, we use the methods of repeated oversampling and interpolation as required to achieve a balanced distribution. That is, we augment samples in all the categories except the majority class, until a balance is obtained. For the interpolation, we use an algorithm called *SMOTE* (Synthetic Minority Oversampling Technique)[22]. This uses a k-nearest neighbours method to first estimate the k samples nearest to a given data sample, after which, between every pair of the given sample and a neighbour, a new feature vector is sampled by linearly interpolating between the feature vector pair. The implementation for this is as provided in the library *imbalanced-learn*[23].

6. RESULTS AND DISCUSSION

For the stroke onset detection task, the commonly used metrics for evaluation are the precision, recall, and f-score, which are calculated by determining the correct and incorrect onset predictions. A correctly detected onset ('hit') is one that lies within a tolerance window (50 *ms* wide) about an unmatched ground truth onset[24]. The average f-score is calculated as the mean value across all the tracks in a dataset. After the hyper-parameter tuning, we obtain a best f-score of 0.972 on the training set, while on the test set, we obtain a comparably high f-score of 0.965.

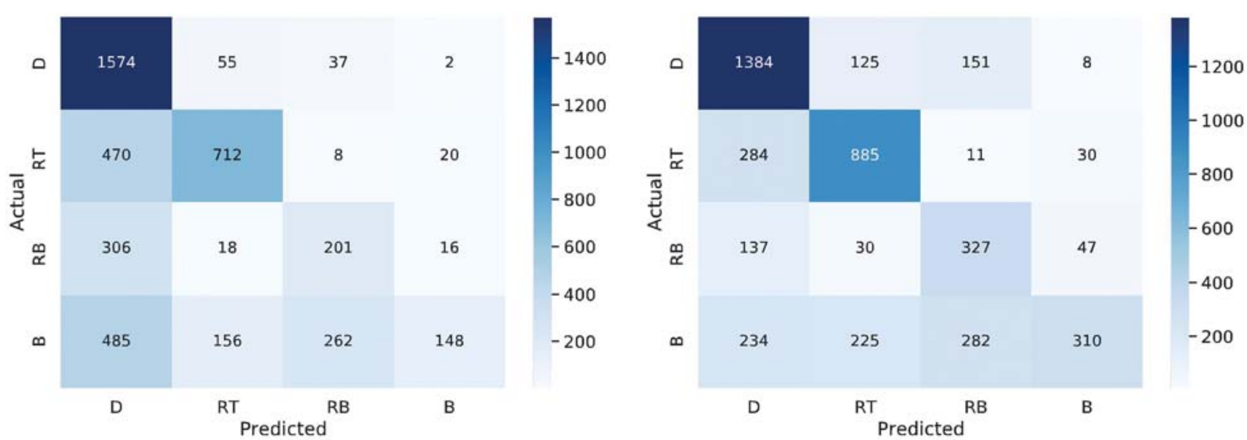
To report stroke classification performance, we use the accuracy as well as average f-score, calculated by averaging the per-class f-score value. We perform cross-validation in a 'Leave-One-Tabla-Out' (LOTO) format, consisting of 3 folds, one each corresponding to each of the 3 tabla sets in the training dataset. This prevents the samples from a given tabla from being present in both the training and evaluation folds and therefore results in a more realistic evaluation of performance. The LOTO-CV results are reported

Table 4. Leave-One-Tabla-Out cross-validation scores (on each fold and the mean) and the test set scores for stroke classification using the various augmentation methods (the highest mean CV and test scores obtained are highlighted in bold).

Method	Cross-validation scores									
	Accuracy				F-score				Test set scores	
	PP	*AS*	*AHK*	Mean	*PP*	*AS*	*AHK*	Mean	Acc.	F-sc.
Baseline (no augmentation)	0.80	0.73	0.57	0.70	0.70	0.58	0.49	0.59	0.59	0.50
Repeated oversampling	0.80	0.74	0.57	0.70	0.70	0.62	0.49	0.60	0.59	0.51
SMOTE	0.79	0.76	0.57	**0.70**	0.70	0.67	0.49	**0.62**	0.59	0.51
Pitch-shifting	0.76	0.66	0.55	0.66	0.69	0.58	0.49	0.59	**0.65**	**0.60**
Pitch-shifting + Rep. Oversampling	0.76	0.67	0.56	0.66	0.69	0.61	0.49	0.60	0.64	0.59
Pitch-shifting + SMOTE	0.76	0.70	0.56	0.67	0.69	0.62	0.48	0.60	0.64	0.58

for each of the 3 tabla sets in Table 4 which also provides the test set performance from a model trained on all the 3 solo tabla training sets. We present different data augmentation and balancing strategies, including some combinations of the two. The baseline method refers to training the classifier on just the originally available dataset. Among the different methods, the use of pitch-shifting leads to a significant rise in the test set score, while resulting in no improvement over the baseline in the average cross-validation f-score. On the other hand, the other two data balancing strategies (repeated oversampling and SMOTE) improve the average CV scores, mainly due to an improvement on the *AS* set alone, while not positively affecting the test set scores (even when performed along with pitch-shifting). This indicates that the latter two methods are perhaps resulting in an overfit on the training data, whereas the pitch-shifting method is helping the model generalise better. Further, among the three training sets, we see a consistently poorer CV performance on the *AHK* set, while the other two are comparable.

From Figure 5, which shows the confusions in stroke label predictions on the test set, we see that the model is clearly biased towards the 'damped' class in the baseline method due to the large class imbalance in the training set. With the use of pitch-shifting augmentation, the errors in all the resonant categories



(a) Base ne (b) t h sh t n

Fig. 5. Confusions in the test set stroke classification with (a) baseline method and (b) pitch-shifting augmentation. (D: 'Damped', RT: 'Resonant Treble', RB: 'Resonant Bass', B: 'Resonant Both')

Table 5. Class-wise validation scores on the AS set in the LOTO-CV. The balanced accuracy is the average per-class accuracy (suitable for imbalanced datasets).

Method	Stroke category								Balanced accuracy
	D		RT		RB		B		
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	
Baseline	0.82	0.93	0.60	0.95	0.86	0.20	0.63	0.29	0.59
Pitch-shifting	0.93	0.69	0.48	0.96	0.71	0.36	0.45	0.39	0.60
Pitch-shifting + SMOTE	0.92	0.75	0.53	0.96	0.74	0.39	0.50	0.43	0.63

reduce, but at the cost of a slight increase in errors in the damped category. Most of the remaining errors are in the resonant bass and resonant both categories. To help explain the previously noted drop in CV accuracies in the course of pitch-shifting based augmentation, we look at the class-wise precision and recall values on the *AS* set alone when it is evaluated in the LOTO-CV procedure (Table 5). Also reported is the 'balanced accuracy' - the average per-class accuracy, which is more suitable for evaluating the performance on imbalanced datasets. The significant drop in the recall value for the damped stroke with the use of pitch-shifting, coupled with its higher count in the dataset, brings out the reason for a drop in the overall accuracy. However, the balanced accuracy improves with the use of pitch-shifting augmentation, and further, with SMOTE. Further, the opposite trends in the precision and recall values for the resonant stroke categories with the use of pitch-shifting explain the lack of improvement in the CV f-score.

In Table 6, we offer a comparison of our baseline cross-validation results (without any augmentation) with a few relevant tabla transcription results from the literature. We also evaluate our method using a random 3-fold CV (as opposed to the LOTO-CV reported in Table 4) to aid this comparison. The dataset size, number of tabla sets, number of target classes, and modes of evaluation are all different in each of the methods. Some of the scores are reported as ranges because the average accuracy reported by Chordia is calculated over a number of individual results on a few subsets of the data. Some of these subsets include single tabla sets out of the total of 7, contributing sometimes to the high average accuracy. Hence, the ranges reported here are an attempt to cover the whole set of accuracies reported under the different conditions. A similar approach was taken to arrive at the accuracy range reported for our method as well. Although a direct comparison is difficult due to the larger target class set in the previous methods (comprising all the distinct bols), a key observation is the huge range of LOTO-CV score reported by Chordia[5], pointing again to the larger problem of generalisation to unseen instruments.

Table 6. Comparing the cross-validation results of our method with previous tabla transcription studies.

Method	**# Target classes**	**# strokes, tabla-sets**	**Random CV acc.**	**LOTO-CV acc.**
Ours	4	6678, 3	0.71 - 0.92	0.57 - 0.80
Sarkar *et al.*[8]	10	3964, 7	0.85	-
Gupta *et al.*[11]	18	8200, 1	0.66	-
Chordia[5]	10-16	16384, 7	0.77 - 0.94	0.15 - 0.95

Next, we look at the results from the feature selection and model hyperparameter tuning. The plot of average LOTO-CV f-score versus the number of features shows a rise till about 30 features, after which it first plateaus and then dips slightly towards the end (Figure 6). The RFE method was then applied to the entire training set in order to pick the best set of 30 features, and the model tuning was performed only using this subset of all features. At the outset, there was no significant improvement in the average CV

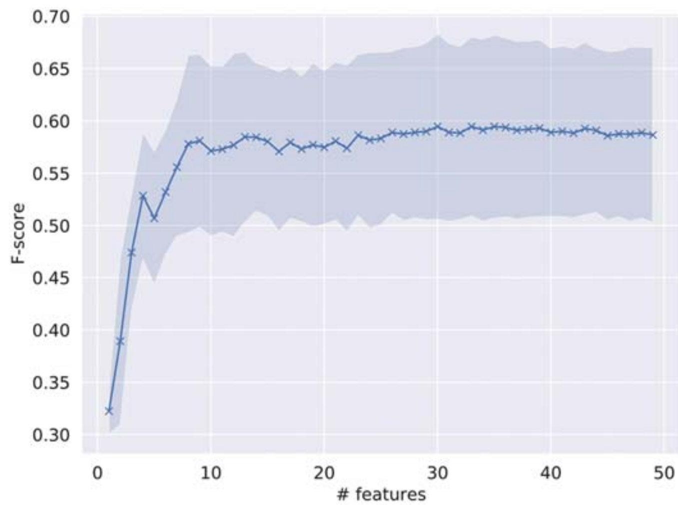


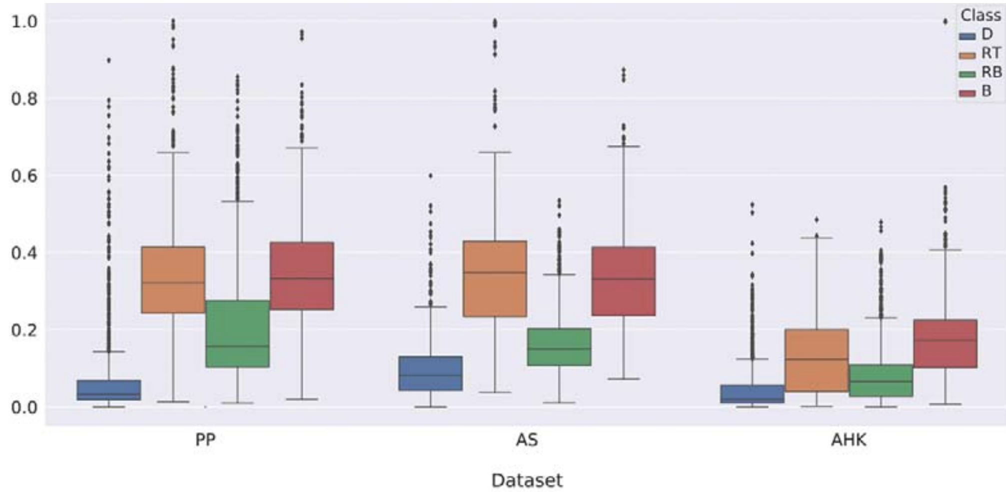
Fig. 6. Plot of average LOTO-CV f-score versus the number of features selected in the recursive feature elimination procedure (RFE). The shaded region represents the standard deviation of the score across the CV folds.

f-score compared to the results of Table 4. A decrease in the number of trees was not found to cause a significant drop in the f-score, perhaps due to the small size of our dataset. And, as expected, at a given number of trees, setting the maximum features to 'None' (*i.e.*, providing each tree with all features) was found to cause a significant drop in performance due to overfitting.

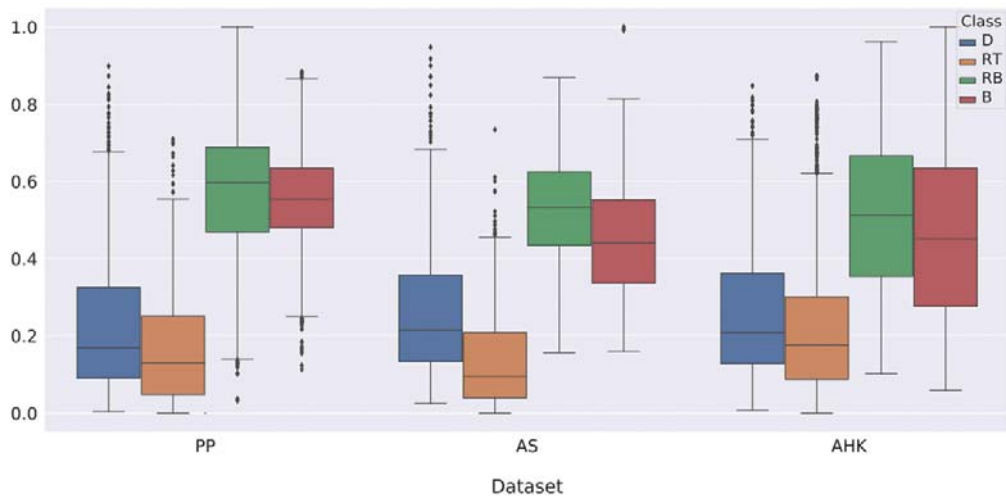
Finally, given the lower performance on the *AHK* set in the LOTO-CV, a closer look at it can help analyse the instrument dependence of the task. As a first step, we look at models fit separately on each of the three subsets in the training set, with each subset augmented using the pitch-shifting method. Table 7 shows the ten most important features according to their contribution to the model fit on each tabla set, in comparison with a model fit on the entire training set. The features highlighted in bold refer to those that are exclusive to a subset. We find that almost all the features are common to the three subsets, although with slightly different rankings. This can be taken as a point in favour of the aptness of the features for

Table 7. The 10 highest ranked features based on importance by a random forest model when fit on different training subsets (entries in bold are features important only to one subset).

Rank	Feature			
	PP	AS	AHK	All
1	Bass onset strength	Bass onset strength	Bass onset strength	Bass onset strength
2	Treble energy sum	MFCC 1	MFCC 12	Treble energy sum
3	Treble onset strength	MFCC 4	Treble energy sum	MFCC 1
4	MFCC 3	MFCC 3	Treble early decay rate	Treble onset strength
5	Treble early decay rate	Treble energy sum	Log attack time	Treble early decay rate
6	MFCC 1	Spec. centroid mean	**Spec. centroid stdev.**	MFCC 2
7	MFCC 2	Treble onset strength	Spec. centroid mean	MFCC 3
8	ZCR mean	Treble early decay rate	MFCC 2	Spec. centroid mean
9	Log attack time	ZCR mean	**Bass early decay rate**	Log attack time
10	MFCC 4	MFCC 12	ZCR mean	ZCR mean



(a) Treble energy sum



(b) Bass onset strength

Fig. 6. Boxplots of the distributions in each training subset of (a) treble energy sum and (b) bass onset strength. (D: 'Damped', RT: 'Resonant Treble', RB: 'Resonant Bass', B: 'Resonant Both')

the stroke classification task, across tabla sets. However, the feature distributions shown in Figure 7, of the top two features (from the last column in Table 7) shed light on the instrument dependence of the feature values. While the distribution of bass onset strength is similar in all the subsets, in the case of treble energy sum, the *PP* and *AS* subsets have a similar distribution that is quite different from that of *AHK*. This is a likely cause of the poor generalisation of the model.

7. CONCLUDING REMARKS

The presented study brings out the challenges in transcribing tabla stroke sequences when labeled training data from the same tabla set is unavailable. Further, the imbalanced nature of the datasets built using recordings of tabla compositions makes it difficult to train an unbiased model, making it necessary to employ appropriate data balancing and augmentation strategies. The use of pitch-shifting augmentation to increase the size of the less populated resonant stroke categories is found to help, but not on all tabla sets - the cross-validation scores do not improve, while the test set scores do. On the other hand, augmenting in the feature space by interpolating between feature vectors improves the cross-validation scores but leads to an overfit when applied to the entire training set, thus not really improving the test set scores. A preliminary analysis of the feature importances in the models trained separately on each tabla set reveals that the same few features are ranked highest in each case. Most of these features are not the standard MFCC or spectral shape-related ones, but happen to be ones that are easier to relate to tabla acoustics (such as onset strength, decay rate, etc). However, the feature distributions vary significantly across tabla sets consistent with the observed instrument-dependence. Therefore, a way to perform more effective data augmentation would be to selectively transform more features that represent the tabla instrument characteristics. The augmentation methods used in the present work are general and do not leverage any specificity of the task. Future work will target acquiring a larger dataset, possibly using unsupervised labeling to reduce the manual transcription effort. Further, data augmentation methods that modify those acoustic attributes that are not important to the distinctions between strokes will be explored to better embed the necessary diversity in the training dataset. Finally, we look forward to a system that can be coupled with musical source separation to achieve the automatic transcription of tabla accompaniment in the typical context of a vocal concert's mixed audio.

8. REFERENCES

[1] Bonnie C. Wade, 1979. Music in India : *The Classical Traditions*. Prentice-Hall.

[2] Ajay Srinivasamurthy, Andre Holzapfel, Kaustuv Kanti Ganguli and Xavier Serra, 2017. "Aspects of Tempo and Rhythmic Elaboration in Hindustani Music: A Corpus Study". In: *Frontiers in Digital Humanities*, **4**, 20.

[3] Martin Clayton, 2020. "Theory and practice of long-form non-isochronous metres". *In: Music Theory Online 26.1*.

[4] Olivier Gillet and Gaël Richard, 2003. "Automatic labelling of tabla signals". *In: Proc. of the 4th Int. Society for Music Information Retrieval Conf.*

[5] Parag Chordia, 2005. "Segmentation and Recognition of Tabla Strokes". *In: Proc. of the 6th Int. Society for Music Information Retrieval Conf.*

[6] Subodh Deolekar and Siby Abraham, 2016. "Classification of Tabla Strokes Using Neural Network". In: *Computational Intelligence in Data Mining* - by Himansu Sekhar Behera and Durga Prasad Mohapatra. Springer India, **1**, 347–356.

[7] Subodh Deolekar and Siby Abraham, 2018. "Tree-based classification of tabla strokes". In: *Current Science* (00113891) 115.9 pp. 1724–1731.

[8] Rajib Sarkar, Ankita Singh, Anjishnu Mondal and Sanjoy Kumar Saha, 2018. "Automatic Extraction and Identification of Bol from Tabla Signal". In: *Advanced Computing and Systems for Security:* by Rituparna Chaki, Agostino Cortesi, Khalid Saeed, and Nabendu Chaki. Springer Singapore, **5**, 139–151.

[9] Shambhavi Shete and Saurabh Deshmukh, 2020. "Analysis and Comparison of Timbral Audio Descriptors with Traditional Audio Descriptors Used in Automatic Tabla Bol Identification of North Indian Classical Music". In: *Proceeding of International Conference on Computational Science and Applications. Ed.* by Subhash Bhalla, Peter Kwan, Mangesh Bedekar, Rashmi Phalnikar, and Sumedha Sirsikar. Springer Singapore, pp. 295–307.

[10] Anirban Patranabis, Kaushik Banerjee, Vishal Midya, Sneha Chakraborty, Shankha Sanyal, Archi Banerjee, Ranjan Sengupta and Dipak Ghosh, 2015. "Harmonic and timbre analysis of tabla strokes". In: *arXiv preprint arXiv:1510.04880*.

[11] Swapnil Gupta, Ajay Srinivasamurthy, Manoj Kumar, Hema A. Murthy and Xavier Serra, 2015. "Discovery of Syllabic Percussion Patterns in Tabla Solo Recordings". In: *Proc. of the 16th Int. Society for Music Information Retrieval Conference.*

[12] Parag Chordia and Alex Rae, 2008. "Tabla Gyan: a System for Realtime Tabla Recognition and Resynthesis". In: *Proc. of the International Computer Music Conference.*

[13] Krish Narang and Preeti Rao, 2017. "Acoustic features for determining goodness of tabla strokes". In: *Proc. of the 18th Int. Society for Music Information Retrieval Conference.*

[14] Rohit M.A. and Preeti Rao, 2018. "Acoustic-Prosodic Features of Tabla Bol Recitation and Correspondence with the Tabla Imitation". In: *Proc. of Interspeech.*

[15] Rohit M.A. and P. Rao, 2019. "Separation of Tabla Strokes from Tabla Solo Audios Using Non-negative Matrix Factorization". In: *Technical Report, I.I.T. Bombay.*

[16] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies and Mark B Sandler, 2005. "A tutorial on onset detection in music signals". In: *IEEE Transactions on Speech and Audio Processing 13.5.*

[17] Bogdanov D., Wack N., Gómez E., Gulati S., Herrera P., Mayor O., Roma G., Salamon J., Zapata J.R. and Serra X., 2013 . "ESSENTIA: an Audio Analysis Library for Music Information Retrieval". In: *Proc. of the 14th Int. Society for Music Information Retrieval Conference.*

[18] Brossier P., Bello J.P. and Plumbley M.D., 2004. "Fast labelling of notes in music signals". In: *Proc. of the 5th Int. Society for Music Information Retrieval Conference.*

[19] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Pretten-hofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M. and Duchesnay E., 2011. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research 12.*

[20] Brian McFee, Colin Raffel, Dawen Liang, Daniel P.W. Ellis, Matt McVicar, Eric Battenberg and Oriol Nieto, 2015. "librosa: Audio and music signal analysis in python". In: *Proc. of the 14th Python in Science Conference.*

[21] David Courtney, 2016. *Manufacture and Repair of Tabla.* Sur Sangeet Services.

[22] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall and W. Philip Kegelmeyer, 2002. "SMOTE: Synthetic Minority Over-sampling Technique". In: *Journal of Artificial Intelligence Research 16.*

[23] Guillaume Lemaître, Fernando Nogueira and Christos K. Aridas, 2017. "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning". In: *Journal of Machine Learning Research 18.17.*

[24] Colin Raffel, Brian McFee, Eric J. Humphrey, Justin Salamon, Oriol Nieto, Dawen Liang and Daniel P.W. Ellis, 2014. "mir_eval: A Transparent Implementation of Common MIR Metrics". In: *Proc. of the 15th Int. Society for Music Information Retrieval Conference.*

The Journal of Acoustical Society of India : Vol. 48, No. 1-2, 2021 (pp. 35-45)

Variation of singing styles within a particular *Gharana* of Hindustani classical music — a nonlinear multifractal study

Archi Banerjee[1,2*], Shankha Sanyal[1,3], Ranjan Sengupta[1] and Dipak Ghosh[1]
[1]Sir C.V. Raman Centre for Physics and Music, Jadavpur University, Kolkata-700032
[2]Rekhi Centre of Excellence for the Science of Happiness,
Indian Institute of Technology Kharagpur, Kharagpur-721302
[3]School of Languages and Linguistics, Jadavpur University, Kolkata-700032
e-mail: archibanerjee7@gmail.com

[Received: 10-11-2020; Revised: 15-12-2020; Accepted: 15-12-2020]

ABSTRACT

Hindustani classical music is entirely based on the "*Raga*" structures. In Hindustani music, a *"Gharana"* or school refers to the adherence of a group of musicians to a particular musical style. *Gharanas* have their basis in the traditional mode of musical training and education. Every *Gharana* has its own distinct features; though within a particular *Gharana*, significant differences in singing styles are observed between generations of performers, which can be ascribed to the individual creativity of that singer. This work aims to study the evolution of singing style among four artists of four consecutive generations from *Patiala Gharana*. For this, *alap* and *bandish* parts of two different *Ragas* sung by the four artists were analyzed with the help of non linear multifractal analysis (MFDFA) technique. The multifractal spectral width obtained from the MFDFA method gives an estimate of the complexity of the signal. The observations from the variation of spectral width give a cue towards the scientific recognition of *Guru-Shisya Parampara* (teacher-student tradition) - a hitherto much-heard philosophical term. From a quantitative approach this study succeeds in analyzing the evolution of singing styles within a particular *Gharana* over generations of artists as well as the effect of globalization in the field of classical music.

1. INTRODUCTION

1.1 A brief introduction to Hindustani classical music

Indian classical music can be compared to the endless sky when its creative aspect is considered. It is believed that in ancient times, there existed a single form of Indian Classical Music. However, the consecutive foreign invasions over the centuries had a great impact on the Indian Classical Music and this created a division into this form of music. This lead to the regeneration of two forms of Indian Classical Music: the Carnatic Music (Usually this style is followed in the Southern parts of India and considered by some as the basic version of Indian Classical Music) and the Hindustani Music (The North Indian and the improvised version of the Indian Classical Music). *Raga* is the heart of Hindustani classical music.

© 2021 Acoustical Society of India

Each *Raga* has a well defined structure consisting of a series of four/five or more musical notes upon which its melody is constructed. However, the way the notes are approached and rendered in musical phrases and the mood they convey are more important in defining a *Raga* than the notes themselves[1]. This leaves ample scope for improvisation within the structured framework of any *Raga*. Every performer of this genre (Hindustani classical music) is essentially a composer as well as an artist. While performing a *Raga* every time, an artist gradually goes on to explore new pathways - new connections between the used notes following his current state of imagination but, at the same time, he maintains the basic discipline of that *Raga*. In this way he slowly moves beyond the strangle hold of the *Raga* grammars and dives into the more subtle and sublime emotional self of the *Raga*.

Alap is the opening section of a *Raga* performance in typical Hindustani classical *(Khayal)* singing/ instrument playing style and acts as the preface of the *Raga*. In the *alap* part the *Raga* is introduced and the paths of its development are revealed using all the notes used in that particular *Raga* and allowed transitions between them with proper distribution over time. *Alap* is usually accompanied by the tanpura drone only and sung at a slow tempo or sometimes without tempo. Then, in case of vocal *Raga (Khayal)* performances, comes the *vilambit bandish* part where the lyrics and rhythmic cycle or *taal* are introduced. *Bandish* is a fixed, melodic composition in Hindustani vocal music, set in a specific *raga*, performed with rhythmic accompaniment by a tabla or pakhawaj, a steady drone, and melodic accompaniment by a sarangi, harmonium *etc*. *Vilambit* is a type of *bandish* which is sung at a very slow tempo, or *laya*, of 10-40 beats per minute. The first paragraph of the song - *Sthayi* is followed by the second one - *Antara*[2]. A *Vilambit bandish* is usually followed by a *Drut* (or higher tempo ~90-120 bpm) *bandish* along with several types of melodic and lyrical improvisations within both the *Vilambit* and *Drut bandishes*.

1.2 *Gharana* tradition of Hindustani classical music and *Guru-Shishya parampara*

Beyond this infinite canvas of individual improvisations, some different styles of presentation are also observed across the country while rendering any particular *raga*. In the context of Hindustani classical music, the major differences in the *Raga* presentation styles are named after different *"Gharanas"*. The coinage *"Gharana"* came from the Hindi word "Ghar" (House). It is commonly observed that the *Gharanas* are named after different places, *viz.*, Agra, Patiyala, Gwalior, Maihar, Bishnupur, Indore etc. The naming of these *Gharanas* mostly indicates towards the origination of these particular musical styles or ideologies. Under the Hindustani Classical Music, the tradition of *"Gharana"*[3, 4] system holds special importance for many listeners. Perhaps, this feature is so unique that no where around the world can one find this sought of a tradition. The *Gharana* system is followed by both the North-Indian as well as the South-Indian forms of Indian classical music. In south India, the term *Gharana* is acknowledged by the word "Sampraya". In ancient times, there existed several Samprayas such as the "Shivmat", the "Bhramamat" and the "Bharatmat"[5]. *Gharanas* have their basis in the traditional mode of musical training and education. Every *Gharana* has its own distinct features; the main area of difference between *Gharanas* being the manner in which the notes are sung. Though, in most cases, even within a particular *Gharana*, significant changes in singing styles are observed between generations of performers. These changes are ascribed to the individual creativity of that singer, which has led to the evolution of the singing style of that particular *Gharana*.

One of the most unique and exclusive feature which is inherent to the teaching process of Indian Classical Music is known as the *"Guru-Shishya"* tradition[6]. History and statistics reveal that majority of the finest artists of Indian Classical Music have been produced through the *Guru-Shishya Parampara* (tradition). In India, the *Gharana* system has contributed to all the three forms of music, that is vocal, instrumental and dance. The *Gharana* comes into existence through the confluence of the *"Guru"* and the *"Shishya"*[7]. A wise *"Guru"* through his intelligence, aptitude and shear practice creates a sense of uniqueness and exclusivity and thereby inculcates a special eminence into his form of music. These attributes and traits are amicably transferred into the talented *"Shishya"* and the particular form of the performing arts thus becomes a tradition. These exceptional qualities are in fact so strong and prominent that the audiences can immediately recognize the *Gharana i.e.* the similarity in singing/dancing/playing style between two artists. It is believed that when so ever the form or style created by the founder *"Guru"* is carried forth till three generations; it turns in to the form of *"Gharana"*. The name of the *Gharana* can be same as the name of the founder *"Guru"*, or can be named after the place where the founder *("Guru")* resided. Throughout the history of the Hindustani classical music tradition, students were often born into a musical family practicing the art of music in the *Guru-Shishya* tradition, which was passed down through hereditary means by the musically gifted members of the family[8]. Traditionally, families belonging to a *Gharana* practiced the art of relaying musical knowledge from one generation to the next, and the music and the particular style of one *Gharana* became the basis for playing, understanding, and critiquing music within the Hindustani classical style. Over time the *Gharana* system has expanded to include a student base that was not necessarily connected through the blood line of the *Guru*. The tradition shifted to accommodate these new students, but only so far as to welcome them into the system that was already in place: the student, though unrelated by blood, still became a part of the *Guru's* family. This practice still remains in certain areas of India, depending on the *Guru*, but with the constant influence of global ideology, the *"Guru-Shishya"* system has had to adapt once more. The students may not be able to live with their *Guru* throughout their lives, but the respect, the devotion and the support offered to the *Guru* remains the same.

1.3 Earlier studies on Gharana tradition and background of the present study

Now-a-days, the "*Gharana*" system in Hindustani classical music is affine to a number of ambiguous ideas among artists. According to a few, *Gharana* system exists pretty distinctively, while some are against this thought. What defines the *Gharana*? Is it really true that the artists of the same *Gharana* keep their unique singing style unchanged over generations or evolution of music takes place like everything else in nature? These questions are still unanswered; at least from a scientific point of view. From the earlier discussions it becomes clear that *Gharana* represents a family of musicians, a well-knit unit evolving, guarding and disseminating the distinctive style through its members, some of whom are well-known performers and some who are not. The wealth of the *Gharana* has always been its knowledge - while performing a *raga* the knowledge of song lyrics, rhythm, ornamentation, which coupled with distinctive voice production, phrasing and sequence, produces a unified whole, an aesthetic form special to each family, a unique lineage of features ascribed to a particular *Gharana*. The literature has very few scientific studies which attempted to analyze the significant musical styles that define a particular *Gharana* [9, 10] or the differences in *Raga* presentation styles observed between different *Gharana*s. Most of the previous studies looked into the aesthetic and prosodic features[11] that distinguish one *Gharana* from another. Datta *et al*.[11] made use of projection pursuit techniques to analyze the similarity between different artists, meant to classify the ragas objectively rather than perceptually. It took the help of various linear features such as MFCC, RMS energy Spectral Irregularity / Centroid to identify the specific features in singers of a particular *Gharana*. This domain of research was still lacking a rigorous scientific study which can objectively try to look for the similarities and dissimilarities in the overall singing style of different artists from the same *Gharana*. To begin our search in such a huge field of study our initial endeavor was to perform an analytical comparison between the renditions of two particular ragas sung by some artists who are said to belong to a particular *Gharana* of Hindustani classical music. In this study, for the first time, we applied robust non-linear tools to extract multifractal features from the complete audio signal waveforms of the *Raga* renditions sung by four generations of singers belonging to a particular *Gharana* (Patiyala). The multifractal techniques used in our study have been discussed in depth in the next section.

1.4 Use of multifractal technique (MFDFA) to identify different singing styles

Previous knowledge suggests that music signals have a complex behavior: at every instant components (in micro and macro scale: pitch, timbre, accent, duration, phrase, melody *etc.*) are closely linked to each other[12, 13]. These properties are peculiar of systems with chaotic, self organized, and generally, nonlinear behavior. Therefore, the analysis of music using linear and deterministic frameworks seems unrealistic and a non-deterministic / chaotic approach is needed in understanding the speech / music signals. Music data is a quantitative record of variations of a particular quality (displacement) over a period of time. One way of analyzing it is to look for the geometric features to help towards categorizing the data in terms of concept[14]. The time evolution of the inherent geometrical structure of a certain music piece can be judged rigorously using latest-state-of-the-art nonlinear technique - fractal analysis, which determines the symmetry scaling behavior of a time series. Fractal analysis of audio signals was first performed by Voss and Clarke[15], who analyzed amplitude spectra of audio signals to find out a characteristic frequency f_c, which separates the white noise from a highly correlated behavior ($\sim 1/f^2$). However, it is well-established experience that naturally evolving geometries and phenomena are rarely characterized by a single scaling ratio; different parts of a system may be scaling differently, *i.e.*, the clustering pattern is not uniform over the whole system. Such a system is better characterized as 'multifractal'[16]. A multifractal can be loosely thought of as an interwoven set constructed from sub-sets with different local fractal dimensions. Real world systems are mostly multifractal in nature. Music too, has non-uniform property in its movement[17, 18] as it is often featured by very irregular dynamics, with sudden and intense bursts of high-frequency fluctuations. Su & Wu[17] showed that both melody and rhythm can be considered as multifractal objects by separating both of them as series of geometric points. Live performances encompass a variety of such musical features including tempo fluctuations[19], notation and timbre variation to name a few. To study such a signal, Multifractal Detrended Fluctuation Analysis (MFDFA) would certainly be a better tool than Detrended Fluctuation Analysis (DFA) as DFA measures the monofractal scaling property of a time series and yields only a single scaling exponent. The MFDFA technique gives us the multifractal spectral width (W) which is a measure of the inherent complexity of the music signal. The MFDFA was first conceived by Kantelhardt *et al.*[20] as a generalization of the standard DFA[21]. We hypothesize that the change of multifractal spectral widths of the signals will give us a cue about the variation and retention of singing styles among the artists belonging to different generations representing the same *Gharana*. Four popular vocalists of four successive generations from a particular *Gharana* (Patiala) of Hindustani music were chosen for this study. The renditions of two well known basic ragas - *Bageshri & Jaijawanti* (containing both *alap* and *vilambit bandish* parts), sung by those four vocalists were analyzed using MFDFA technique and the spectral width values corresponding to different parts of the *Raga* renditions were compared for the four artists to study the evolution in their singing styles.

2. EXPERIMENTAL DETAILS

2.1 Choice of Music Signals

In this work, our objective was to study the similarity & changes in the singing pattern of a particular *raga* over generations of artists of the same *Gharana*. For the experiments, recordings of the renditions of two ragas *(Bageshri & Jaijawanti)*, containing the introductory *alap* section as well as both the *sthayi* and *antara* parts of the low tempo *vilambit bandish*, were collected from four artists of consecutive generations belonging to a particular *Gharana (Patiyala)* of Hindustani classical music. For each *raga* the chosen *bandish* was same for all the four artists.

2.2 Processing of Music Signals

All the signals were digitized at the rate of 44100 samples / sec 16 bit format. The *alap* part, *sthayi* and *antara* of the *bandish* part were cut out separately from each rendition before detailed analysis. It is expected that in the *alap* part, note combinations or improvisations will differ for different vocalists while establishing the *raga* and hence there was significant variation in the length of the *alap* pieces chosen for our analysis. So, to minimize the variation due to improvisations, in case of each vocalist, about 20 seconds of the *alap* part were cut out which led only to identification of the *raga*. The said 20 seconds clips were selected by an eminent musician with more than 20 years of experience in performing Hindustani classical music. On the contrary, the *bandish* part has lesser chances of variation in note combinations as for a particular *raga*, the same *bandish* was sung by all the vocalists keeping the melody structure almost same. Although for different vocalists significant variations in the scansion of the *bandish i.e.* the distribution of the lyrics over the whole rhythm cycle of the taal are expected. These clips were then selected for analysis.

3. METHODOLOGY

For the analysis each of the chosen *alap* parts of 20 seconds duration was divided into 4 equal parts of 5 seconds and their multifractal spectral widths were calculated using the MFDFA technique. The variation in the spectral widths among the 4 vocalists of 4 different generations was observed separately for the 2 chosen ragas. Similar observations were found in case of the *sthayi* and *antara* part of the *vilambit bandishes*. The detailed algorithm for MFDFA technique is given in Section 3.1.

3.1 Multifractal Detrended Fluctuation Analysis of sound signals

The time series data obtained from the sound signals are analyzed using MATLAB and for each step an equivalent mathematical representation is given which is taken from the prescription of Kantelhardt *et al.*[20]. In MFDFA technique, first the noise-like structures of the signal was converted into a random walk like signals. It can be represented as:

$$Y(i) = \sum_{k=1}^{i} \left(x_k - \overline{x} \right) \tag{1}$$

where $\overline{x}$ is the mean value of the signal.

The integration reduced the level of noise present in experimental records and finite data. Then the whole length of the signal is divided into Ns no of non-overlapping 'intervals' (int) consisting of certain no. of samples. For *'s'* as sample size and *'N'* the total length of the signal, the intervals are:

$$N_s = int\left(\frac{N}{s}\right) \tag{2}$$

Since the length N of the series is often not a multiple of the considered time scale s, a short part at the end of the profile may remain. In order not to disregard this part of the series, the same procedure is repeated starting from the opposite end. There by, 2 N_s segments are obtained altogether. The local RMS variation for any sample size *'s'* is the function $F(s, \nu)$. This function can be written as follows:

$$F^2(s, \nu) = \frac{1}{s} \sum_{i=1}^{s} \left\{ Y\left[(\nu-1)s+i\right] - y_\nu(i) \right\}^2 \tag{3}$$

For $\nu = N_s + 1$........, 2 N_s, where y_ν (i) is the least square fitted value in the bin . In this work, a least square linear fit using first order polynomial (MF-DFA -1) is performed. The q-order overall RMS variation for various scale sizes can be obtained by the use of following equation

$$F_q(s) = \left\{ \frac{1}{N_s} \sum_{\nu=1}^{N_s} [F^2(s,\nu)]^{\frac{q}{2}} \right\}^{\left(\frac{1}{q}\right)} \tag{4}$$

where q is an index that can take all possible values except zero, because in that case the factor $1/q$ is infinite. The scaling behavior of the fluctuation function is obtained by drawing the log-log plot of $F_q(s)$ vs. s for each value of q.

$$F_q(s) \sim s^{h(q)} \tag{5}$$

In general, the exponent $h(q)$ may depend on q. For stationary time series, $h(2)$ is identical to the well-known Hurst exponent *'H'*. Thus, we will call the function $h(q)$ generalized Hurst exponent (Fig. 1a). The Hurst exponent is measure of self-similarity and correlation properties of time series produced by fractal. The presence or absence of long range correlation can be determined using Hurst exponent. A monofractal time series is characterized by unique $h(q)$ for all values of q. The generalized Hurst exponent $h(q)$ of MF-DFA is related to the classical scaling exponent $\tau(q)$ by the relation

$$\tau(q) = qh(q) - 1 \tag{6}$$

A monofractal series with long range correlation is characterized by linearly dependent q order exponent $\tau(q)$ with a single Hurst exponent *'H'* ($h(q)$ for q=2). Multifractal signal on the other hand, possess multiple Hurst exponent and in this case, $\tau(q)$ depends non-linearly on q[22]. Another way to characterize a multifractal series is the singularity spectrum $f(\alpha)$, which is related to $\tau(q)$ via a Legendre transform[23, 24] as follows :

$$\alpha = \tau'(q) \text{ and } f(\alpha) = q\,\alpha(q) - \tau(q) \tag{7}$$

Here, α is the singularity strength or Hölder exponent, while $f(\alpha)$ denotes the dimension of the subset of the series that is characterized by α. Using Eq. (6), α and $f(\alpha)$ can be directly related to $h(q)$

$$\begin{aligned} a(q) &= h(q) + qh'(q) \\ f(a) &= q[a(q) - h(q)] + 1 \end{aligned} \tag{8}$$

Here, $h'(q)$ denotes the first derivative of $h(q)$ with respect to q. The $f(\alpha)$ vs α plot or the multifractal spectrum (Fig. 1b) is capable of providing information about relative importance of various fractal exponents in the series *e.g.,* the width of the spectrum denotes range of exponents. A quantitative characterization of the spectra may be obtained by least square fitting it to a quadratic function[25] around the position of maximum α_0,

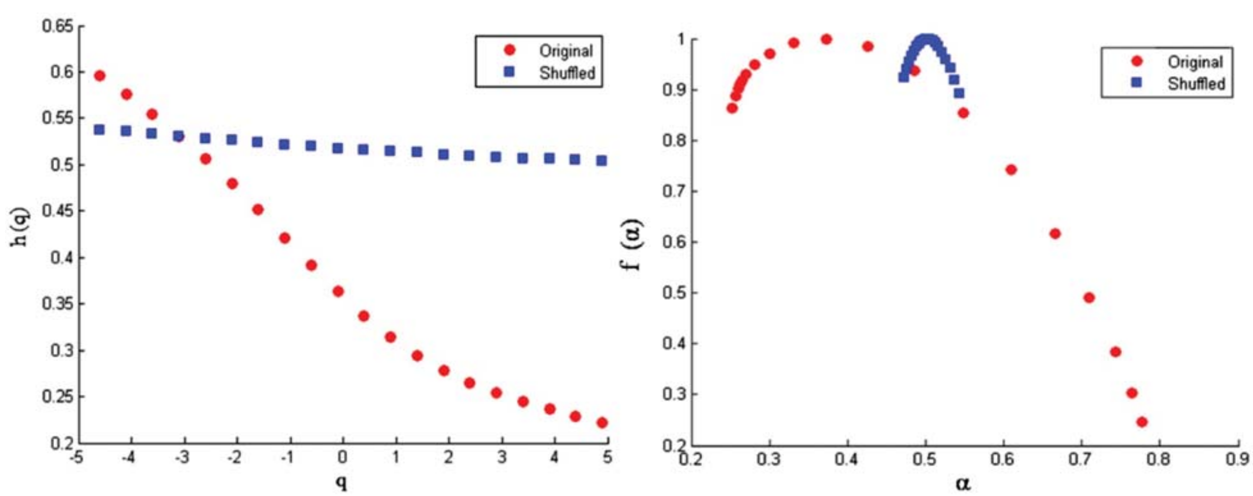
$$f(a) = A(a - a_0)^2 + B(a - a_0) + C \tag{9}$$

where C is an additive constant $C = f(\alpha_0) = 1$. B indicates the asymmetry of the spectrum. It is zero for a symmetric spectrum. The width of the spectrum can be obtained by extrapolating the fitted curve to zero. Width W is defined as,

$$W = a_1 - a_2, \text{ with } f(\alpha_1) = f(\alpha_2) = 0 \tag{10}$$

The width of the spectrum gives a measure of the multifractality of the spectrum. Greater is the value of the width W greater will be the multifractality of the spectrum. For a monofractal time series, the width will be zero as $h(q)$ is independent of q.

The origin of multifractality in a sound signal time series can be verified by randomly shuffling the original time series data[26]. In general, two different types of multifractality are present in a time series data: (i) Multifractality due to a broad probability density function for the values of the time series. Here,



(a) (b)

Fig. 1. (a) Sample $h(q)$ vs q plot and (b) sample $f(\alpha)$ vs α plot for original and shuffled series corresponding to a specific experimental audio signal.

the multifractality of the time series cannot be removed by random shuffling and the shuffled data has the same variation of *h(q)* as the original data (ii) Multifractality due to a variety of long-range correlations due to the small and large fluctuations. In this case, the probability density function of the values can be a regular distribution with finite moments, for *e.g.* a Gaussian distribution. The corresponding shuffled series will exhibit non-multifractal scaling, since all long-range correlations are destroyed by the shuffling procedure. All long range correlations that existed in the original data are removed by this random shuffling and what remains is a totally uncorrelated sequence. Hence, if the multifractality of the original data was due to long range correlation, the shuffled data will show non-fractal scaling. If any series has multifractality both due to long range correlation as well as due to probability density function, then the shuffled series will have smaller width W and hence weaker multifractality than the original time series as is evident from Fig. 1.

4. RESULTS AND DISCUSSION

The multifractal spectral width *(W)* was computed for each part of the *raga* clips for all the four artists. Higher the value of *W*, higher is the degree of complexity present in the signal. Thus change in the spectral width due to change of note combinations, transitions between the notes; tempo and rhythm variations in a musical piece would be significantly important to characterize the singing style of different artists. Table 1 shows all the spectral width values where Artist 1 represents the oldest among the 4 artists of *Patiyala Gharana* that we chose to study. Artists 2, 3 and 4 represent singers of consecutive generations from the same *Gharana.*

Table 1. Variation of the spectral width values for all 4 artists while rendering 2 chosen ragas

Raga used			**Artist 1**	**Shuffled width**	**Artist 2**	**Shuffled width**	**Artist 3**	**Shuffled width**	**Artist 4**	**Shuffled width**
Bageshri	*Alap*	Part 1	0.49	0.11	0.73	0.02	0.49	0.07	0.63	0.01
		Part 2	0.63	0.09	0.72	0.03	0.42	0.03	0.63	0.03
		Part 3	0.77	0.07	0.86	0.01	0.54	0.09	0.53	0.03
		Part 4	0.46	0.11	0.75	0.02	0.65	0.07	0.83	0.06
	Bandish Sthayi	Part 1	0.37	0.09	0.82	0.02	0.72	0.06	0.41	0.03
		Part 2	0.65	0.08	0.65	0.05	0.62	0.05	0.62	0.01
		Part 3	0.45	0.08	0.77	0.07	0.59	0.11	0.41	0.03
		Part 4	0.48	0.10	0.66	0.06	0.65	0.09	0.34	0.03
	Bandish Antara	Part 1	0.28	0.13	0.49	0.03	0.47	0.09	0.48	0.02
		Part 2	0.55	0.09	0.67	0.04	0.54	0.05	0.55	0.02
		Part 3	0.59	0.09	0.66	0.04	0.56	0.10	0.46	0.03
		Part 4	0.46	0.16	0.49	0.07	0.69	0.06	0.41	0.01
Jaijawanti	*Alap*	Part 1	0.57	0.08	0.87	0.08	0.30	0.02	0.60	0.04
		Part 2	0.91	0.25	0.84	0.04	0.63	0.01	0.64	0.02
		Part 3	0.72	0.07	0.57	0.07	0.55	0.08	0.61	0.04
		Part 4	0.90	0.09	0.86	0.04	0.56	0.04	0.56	0.02
	Bandish Sthayi	Part 1	0.79	0.08	0.78	0.05	0.27	0.07	0.56	0.02
		Part 2	0.90	0.03	0.92	0.05	0.37	0.05	0.52	0.02
		Part 3	0.79	0.07	0.62	0.03	0.36	0.05	0.37	0.01
		Part 4	0.72	0.06	0.58	0.02	0.26	0.07	0.40	0.05
	Bandish Antara	Part 1	0.58	0.03	0.77	0.05	0.30	0.05	0.37	0.02
		Part 2	0.56	0.04	0.88	0.02	0.52	0.03	0.29	0.01
		Part 3	0.61	0.05	0.50	0.03	0.15	0.08	0.41	0.04
		Part 4	0.59	0.01	0.71	0.05	0.37	0.09	0.47	0.03

From Table 1 it is evident that the average multifractal spectral width is higher in case of *Raga Jaijawanti* than that of *Raga Bageshri* for all 4 artists both in the *alap* part as well as the entire *bandish* part. Thus we can conclude that the overall complexity of *Raga Jaijawanti* is higher than that of *Raga Bageshri*. To make the trends in our data easier to visualize, variation of multifractal spectral widths among the four generations of artists from *Patiyala Gharana* while singing different parts of the renderings of *Raga Bageshri* and *Raga Jaijawanti* were plotted separately in Fig. 2. The following figures - Fig. 2a and Fig. 2b represent the variation of multifractal spectral width *(W)* for the 4 artists while rendering *alap* parts of *Raga Bageshri* and *Raga Jaijawanti* respectively. Fig. 2c and Fig. 2d represent the same for the *Bandish Sthayi* parts whereas Fig. 2e and Fig. 2f represent the same for *Bandish Antara* parts corresponding to the chosen two *Ragas*. The error bars given in all the following figures represent the computational errors introduced in the multifractal algorithm used in this work.

The following observations can be drawn from a careful study of Figures 2(a-f):

1. Comparing Fig. 2a, Fig. 2c, Fig. 2e with Fig. 2b, Fig. 2d, Fig. 2f, it can be easily observed that spectral width variation among the 4 artists of the same *Gharana* is higher in case of *Raga Jaijawanti* than *Raga Bageshri*. This complexity variation in case of *Raga Jaijawanti* is more prominent during *bandish sthayi* and *antara* parts compared to the *alap* part. These observations may be interpreted as following: In case of *Raga Bageshri* the singing style of the older generation artist is maintained more strictly by his successors than in case of *Raga Jaijawanti* where the new generation artists incorporated their own styles more frequently.
2. The complexity variation among the 4 artists is least in case of *Bageshri antara* (Fig. 2e) while largest in case of *Jaijawanti antara* (Fig. 2f).
3. In general for both *Ragas*, Artist 2 features greater average multifractal spectral width than other three artists both in *alap* part as well as *bandish* part.
4. Analysing the renderings of *Raga Bageshri* sung by the four artists, we can observe prominent similarities between Artist 1 and Artist 2 in the complexity variation pattern among the 4 parts of the *alap* section, but the artists of the newer generations (Artist 3 and Artist 4) differed from them following their own ways. In the *bandish sthayi* part, significant similarity was observed between Artist 2 and Artist 4, whereas Artist 1 and Artist 3 differed from them in a similar manner. In the *antara* part of the *Bageshri bandish*, a great degree of similarity was observed between the singing styles of Artist 1, Artist 2 and Artist 4, but Artist 3 slightly differed from all three of them. Though in *Raga Bagesgri*, the variation in spectral width among the four generations of artists is not very pronounced, but both in the *alap* section and the entire *bandish* section, Artist 1, Artist 3 and Artist 4 (*i.e.*, the artists of first, third and fourth generation respectively) feature lower absolute values of multifractal spectral width compared to Artist 2 who is representing the second generation of artists from the *Patiyala Gharana*.
5. In case of *Raga Jaijawanti*, striking similarity is observed between Artist 1 and Artist 3 in the complexity variation pattern among the 4 parts of the *alap* as well as the *bandish sthayi* part whereas in *antara* part of the *bandish* Artist 3 resemble more with Artist 2 and Artist 4 resemble more with Artist 1, though the absolute values of the multifractal spectral width are much lesser for Artist 3 and Artist 4 compared to Artist 1 and Artist 2 in both *alap* and *sthayi* and *antara* parts of the *bandish*. So, it is evident that while singing *Raga Jaijawanti* the artists of the younger generations are trying to incorporate the singing style of the artists from the older generations as per their choice.
6. In *alap* part of both the *Ragas* Artist 4 significantly differs from other three artists, but in the *Bandish* (both in the *sthayi* and *antara* parts) he shows resemblance with other artists.
7. In most of the time segments we get varying complexity which has a clear tendency to increase from Artist 1 to Artist 2 but mostly decrease in case of the contemporary artists (Artist 3 and Artist 4 in our case).

Thus, the multifractal analysis of music signals can be efficiently used in analyzing the singing styles of different artists while performing any *Raga*.

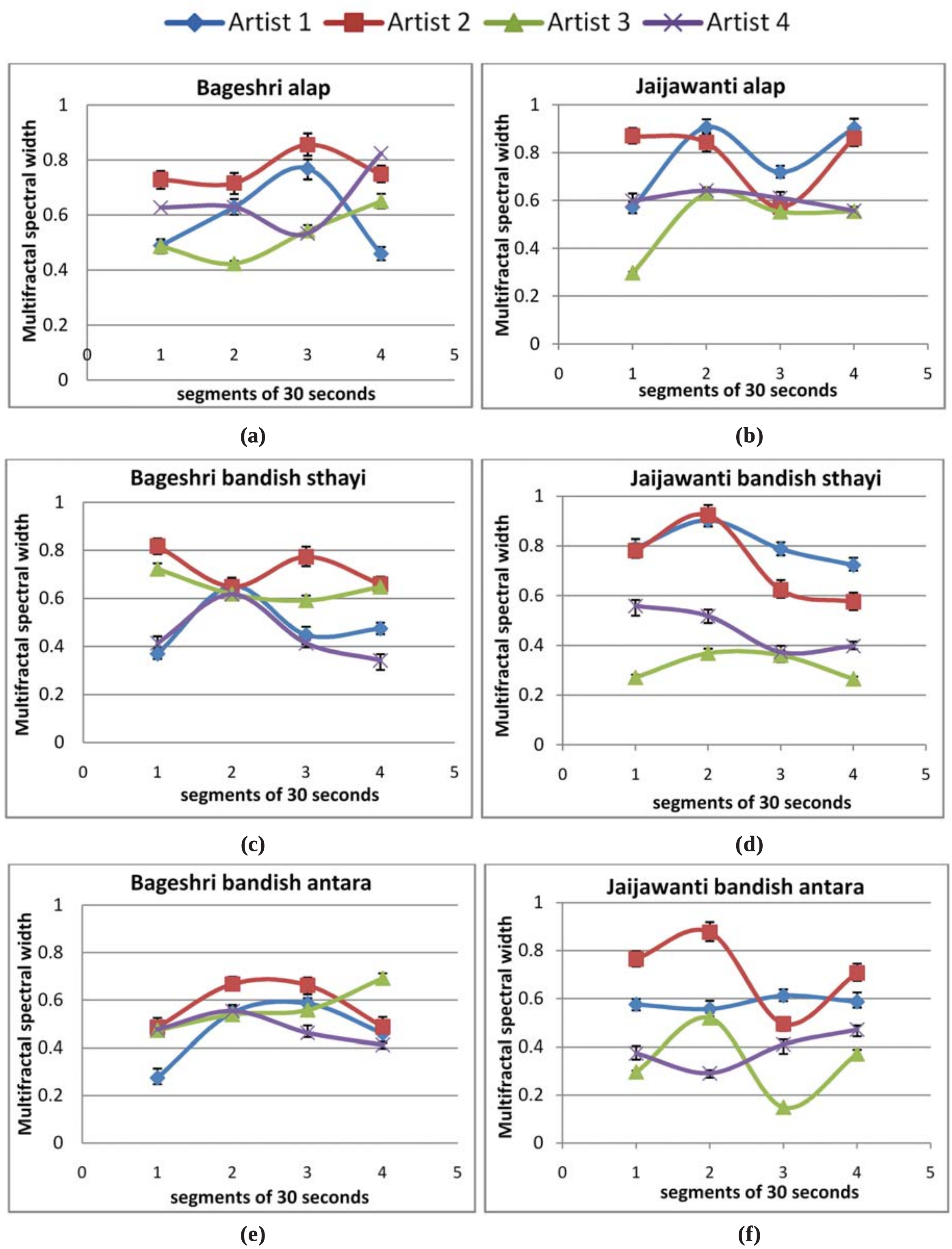


Fig. 2. Variation of multifractal spectral widths in the (a, b) *Alap* parts; (c, d) Bandish shtayi parts and (e, f) *Bandish antara* parts of *Raga Bageshri* and *Raga Jaijawanti* respectively for the chosen four artists representing the four generations of artists from *Patiyala Gharana*.

5. CONCLUSION

Summarizing all the observations obtained from the multifractal detrended fluctuation analysis of the acoustical waveforms of the renderings in *Raga Bageshri* and *Raga Jaijawanti* by four artists of four consecutive generations belonging to the *Patiyala Gharana* of Hindustani classical music we can conclude that Artist 2 (who represents the second generation among the chosen four vocalists) seems to be completely different from the other three artists in *bandish* as well as *alap* parts in terms of complexity of the signal. Artist 3 and Artist 4, who represent the third and fourth generations of artists from *Patiyala Gharana* respectively, resemble Artist 1 of the oldest or first generation in specific time segments. From this trend we may predict that the contemporary artists Artist 3 and Artist 4 are trying to incorporate the style of Artist 1 in their singing sometimes. The variation in multifractal spectral width in different parts of the *alap* or *bandish* section of any *Raga* rendering reveals that in most of the time segments we get varying complexity which has a clear tendency to increase from Artist 1 to Artist 2 but mixed response or more commonly a decrement in the spectral width values in case of the contemporary artists. Analysis by this process using nonlinear chaos based multifractal techniques on the acoustical waveforms of different *Raga* renderings therefore yielded comparison of the singing styles among four vocalists of consecutive generations from a particular *gharana*, which serves our interest. Among all the chosen four artists, Artist 2 is both a direct disciple as well as a close blood relation family member of Artist 1. Artist 3 had the opportunity to learn *Raga* music directly from both Artist 1 and Artist 2 at different times. Artist 4 is a direct disciple of Artist 3. Now, on one hand, the similarities in the variation pattern of the multifractal spectral width (or simply the nonlinear acoustic complexity) in the *alap* and *bandish* sections of the renderings of *Raga Bageshri* and *Raga Jaijawanti* give us hint towards the preservation of some specific *Raga* rendering styles or the manifestation of *Guru-Shishya* tradition within a specific *Gharana* of Hindustani classical music. On the other hand, the dissimilarities in the singing pattern among these four artists indicate towards the improvisational tendencies of the individual artists which may also be affected by the ongoing globalization in every field of our daily lives. This is only a pilot study. Further more, our objective can be attained by the detailed analysis of the extracted notes used in the *Raga* renderings and corresponding note to note transitions *etc.* Also, in future, the whole *alap* part as well as different other parts of several complete *Raga* performances should be analyzed and same study should be done for other *Gharanas* of Hindustani classical music to reach a more convincing conclusion.

6. ACKNOWLEDGEMENT

Archi Banerjee acknowledges the Department of Science and Technology (DST), Govt. of India for providing the DST CSRI Post Doctoral Fellowship (DST/CSRI/PDF-34/2018) to pursue this research work. Shankha Sanyal acknowledges DST CSRI, Govt of India for providing the funds related to this Major Research Project (DST/CSRI/2018/78 (G)) and the Acoustical Society of America (ASA) for providing the International Students Grant.

7. REFERENCES

[1] L.L.M.P. Sheba and A. Revathy, 2014, April. *Raga* identification using clustering algorithm. *In 2014 International Conference on Communication and Signal Processing. IEEE.* pp. 1932-1936.

[2] S. Sanyal, A. Banerjee, A. Patranabis, K. Banerjee, R. Sengupta and D. Ghosh, 2016. A study on Improvisation in a Musical performance using Multifractal Detrended Cross Correlation Analysis. *Physica A: Statistical Mechanics and its Applications,* **462**, 67-83.

[3] A. Brahaspati, 2002. *Sangeet Ratnakar, Sangeet Karyalay, Hathras.*

[4] S. Mahabharati, 2011. The Oxford Encyclopaedia of the Music of India. *Oxford University Press.*

[5] P.S.S. Bodh, 1992. *Madya Pradesh Hindi Granth Academy*.

[6] G. Hothi, 2013. An Interrogative Analysis of the Imdadkhani *Gharana* of Hindustani Classical Music. *Indian Journal of Arts*, **1**(1), 10-12.

[7] S.K. Chaubey, 1977. Sangeet ke *Gharano* ki Charcha, *Uttar Pradesh Hindi Granth Academy, Lucknow*.

[8] R.B. Qureshi, 1987. Musical sound and contextual input: A performance model for musical analysis. *Ethnomusicology,* **31**(1), 56-86.

[9] B. Silver and R.A. Burghardt, 1976. On becoming an Ustād: Six life sketches in the evolution of a Gharānā. *Asian Music,* pp. 27-58.

[10] M. Wright and D. Wessel, 1998. An improvisation environment for generating rhythmic structures based on north indian" Tal" patterns. In ICMC.

[11] A.K. Datta, S.S. Solanki, R. Sengupta, S. Chakraborty, K. Mahto and A. Patranabis, 2017. Gharana Identification. *In Signal Analysis of Hindustani Classical Music. Springer, Singapore.* pp. 125-141

[12] M. Bigerelle and A. Iost, 2000. Fractal dimension and classification of music. *Chaos, Solitons & Fractals,* **11**(14), 2179-2192.

[13] R. Sengupta, N. Dey, A.K. Datta and D. Ghosh, 2005. Assessment of musical quality of tanpura by fractal-dimensional analysis. *Fractals,* **13**(03), 245-252.

[14] R.L. Devaney, L. Devaney and L. Devaney, 1989. An introduction to chaotic dynamical systems, vol. **13046**. Reading: Addison-Wesley.

[15] R.F. Voss and J. Clarke, 1975. '1/fnoise'in music and speech. *Nature,* **258**(5533), 317-318.

[16] R. Lopes and N. Betrouni, 2009. Fractal and multifractal analysis: a review. *Medical image analysis,* **13**(4), 634-649.

[17] Z.Y. Su and T. Wu, 2006. Multifractal analyses of music sequences. *Physica D: Nonlinear Phenomena,* **221**(2), 188-194.

[18] L. Telesca and M. Lovallo, 2011. Revealing competitive behaviours in music by means of the multifractal detrended fluctuation analysis: application to Bach's Sinfonias. Proceedings of the Royal Society A: Mathematical, *Physical and Engineering Sciences,* **467**(2134), 3022-3032.

[19] J.G. Holden, G.C. Van Orden and M.T. Turvey, 2009. Dispersion of response times reveals cognitive dynamics. *Psychological review,* **116**(2), 318.

[20] J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde and H.E. Stanley, 2002. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications,* **316**(1-4), 87-114.

[21] C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley and A.L. Goldberger, 1994. Mosaic organization of DNA nucleotides. *Physical review e,* **49**(2), 1685.

[22] Y. Ashkenazy, D.R. Baker, H. Gildor and S. Havlin, 2003. Nonlinearity and multifractality of climate change in the past 420,000 years. *Geophysical research letters,* **30**(22).

[23] J. Feder, 1988. Random walks and fractals. *Fractals. Springer, Boston, MA*. pp. 163-183.

[24] H.O. Peitgen, H. Jürgens and D. Saupe, 2006. Chaos and fractals: new frontiers of science. *Springer Science & Business Media*.

[25] Y.U. Shimizu, S. Thurner and K. Ehrenberger, 2002. Multifractal spectra as a measure of complexity in human posture. *Fractals,* **10**(01), 103-116.

[26] A. Figliola, E. Serrano, J.A.P. Rostas, M. Hunter and O.A. Rosso, 2007. May. Study of EEG brain maturation signals with multifractal detrended fluctuation analysis. *In AIP Conference Proceedings. American Institute of Physics*. **913**(1), 190-195.

The Journal of Acoustical Society of India : Vol. 48, No. 1-2, 2021 (pp. 46-55)

Probabilistic analysis of three *Ragas* of the Bhairavi *Thaat*

Apra Lipi and Soubhik Chakraborty
Department of Mathematics, Birla Institute of Technology, Mesra
Ranchi-835215, Jharkhand, India
e-mail: soubhikc@yahoo.co.in, apralipi.17@gmail.com

[Received: 19-11-2020; Revised: 10-12-2020; Accepted: 10-12-2020]

ABSTRACT

The present work provides a probabilistic analysis of three different ragas of the same thaat. The thaat taken for our study is Bhairavi. The ragas selected from the thaat Bhairavi are:

- Raga Bhairavi (which is an individual raga and should not be confused with the thaat Bhairavi)
- Raga Bilaskhani Todi
- Raga Malkauns

The analysis covers the conditional and unconditional probabilistic behaviour of the three different ragas. Further the study includes the comparative study of aaroh-avroh (ascent-descent sequential pattern) of the ragas and verifying whether the respective bandishes (song like raga compositions) are following the aaroh-avroh pattern. And finally, a comparative study of the strings (note assemblies) used in the bandishes of different ragas is made. Our objective is to provide a mathematical approach for different ragas of the same thaat by analysing their probabilities. The thaat is kept same so as to have common notes for ready comparison. The experimental results are encouraging.

1. INTRODUCTION

Music is the craft of sound as expected and coordinated to the standards of pitch, musicality, and concordance. North Indian classical music (NICM), or Hindustani music, is an old melodic type of India that arose out of a social amalgamation of the Vedic serenade convention and customary Persian music[1]. The main thought in this arrangement of music are ragas (which is also true for South Indian classical music (SICM) or Carnatic music), which are depicted as melodic compositions fit for initiating explicit dispositions or feelings.

The ordinance of standard NICM ragas is sorted and coordinated around a progression of heptatonic scales known as thaats. In the most generally acknowledged NICM framework, there are 10 thaats comprising of various successive mixes of 12 notes. They speak to ten heptatonic scale families which are utilized to order the standard of North Indian Classical ragas into apparently comparative gatherings. Not all ragas in a given thaat incorporate each note of that parent thaat, yet all ragas in a thaat can be gotten from its characterizing scale.

© 2021 Acoustical Society of India

Thaats provide a useful classification framework, but the core of NICM is the raga[2]. A raga is a melodic structure with fixed notes and a set of rules that characterize a particular mood which is conveyed through performance. A raga has a typical aroh-avroh (ascent-descent allowable sequence) pattern. The way a particular note or a note combination is to be rendered is typical of the raga. Interestingly, despite these binding rules, the artist has infinite freedom to express himself / herself within the gamut of the raga. The rules actually help in characterizing the specific raga mood[3].

Each raga uses a set of five or more notes from the seven comprising its parent thaat to construct a melody. Multiple ragas are generated from a single thaat, each distinguished by its own signature phrase (pakad) and a defined frequency of occurrence of particular notes, vadi being the most prominent note and samvadi being the second most prominent[4]. This feature allows two ragas to have the exact same note selection, yet sound different due to varying emphasis on the notes[2].

The ten thaats widely accepted are: Asavari, Bhairav, Bhairavi, Bilawal, Kafi, Kalyan, Khamaj, Marwa, Poorvi and Todi. Here we have taken thaat Bhairavi and three of its ragas for analysis. Thaat Bhairavi is one of the ten basic thaats of Hindustani music from the Indian subcontinent. It is also the name of a raga within this thaat. Ragas in Bhairavi thaat include:

- Bhairavi
- Bilaskhani Todi
- Bhupal Todi
- Kaunsi Kanada
- Komal Rishabh Asavari
- Malkauns

Here in this study we have taken three ragas *i.e.*, raga Bhairavi, raga Bilaskhani Todi and raga Malkauns.

Bhairavi raga is named after the Shakti or feminine aspect of the enormous life power, which is represented as a partner of Shiva (Bhairava). It is commonly accepted that there are no fixed presentation rules for Bhairavi, and that it is left to the creative mind and aptitude of the craftsman to make designs that are tastefully satisfying. In spite of the fact that the performer has the opportunity to present expressions of other raga, Bhairavi has such an unmistakable mind-set and such trademark melodic examples that a prepared audience can quickly remember it. Contingent on the melodic and beautiful substance of the tunes, Bhairavi can have shades of a few passionate articulations going from sentimental and tempting to reverential; however it is generally appropriate for communicating the misery and agony of partition. Raga Bilaskhani Todi and Raga Asavari (with flat Re) have a similar tone material as Bhairavi, and even some normal melodic developments. Notwithstanding, the extra tones that are so ordinarily allowed in Bhairavi are carefully prohibited in these ragas[5]. Bhairavi is often called a Sadabahar raga, *i.e.*, sounds melodious if rendered properly anytime throughout the day. In fact, this was a nice reason to pick up Bhairavi thaat for our study. The choice enabled us to include raga Bhairavi. We do not intend to get into the controversial time theory in Hindustani ragas in the present paper but instead focus on the relative probabilities of notes, both unconditional and conditional.

Remark: One should not confuse between raga Bhairavi which is an individual raga and thaat Bhairavi which is a raga group according to scale (with Re-Ga-Dha-Ni all komal).

Bilaskhani todi is supposedly a creation of Bilas Khan, son of the legendary vocalist Miya Tansen whose name is also associated with the raga Miya ki Todi[2]. It is believed that Bilas Khan, the son of Tansen, created this Raga when Tansen was on his deathbed. As the name suggests raga Bilaskhani Todi is a type of Todi but interestingly enough, it does not belong to Todi thaat. The reason is that Bilaskhani Todi has the notes Re, Ga, Dha, Ni komal (flattened form) and due to this the notes are same as that of raga Bhairavi. This is why raga Bilaskhani Todi is related to thaat Bhairavi. Although it has the same scale as modern Bhairavi, it has the melodic characteristics of Todi, that is, a strong flat Dha (on which the ascent often begins) and a strong (very) flat Ga, and the typical meandering around Re and Ga[6].

Malkauns is a serious, meditative raga, and is developed mostly in the lower octave (mandra saptak) and in a slow tempo (vilambit laya). 'Heavier' ornaments such as meend, gamak and andolan are used rather than 'lighter' ornaments such as murki and khatka. Komal Ni is generally considered the starting note (graham swara), and the notes komal Ga and komal Dha are performed with vibrato (andolit). All five swaras can function as pausing notes. The best time for this raga is late night. The effect of the raga is soothing and intoxicating.

Remarks

1. Elements of Malkauns can be combined with other ragas, resulting in melodies such as Kaunsi Kanada, Kaunsi Bhairav, Jogkauns and several others.

   Chandrakauns is yet another raga which has the same notes as Malkauns except for Ni, which is natural. The prominence of natural Ni makes this raga distinct from Malkauns. Ragas like Harikauns, Madhukauns, Nandkauns & Sundarkauns are also modern creations and share the suffix-kauns. However, their similarity with Malkauns is limited. Sampurna Malkauns is an old raga; also includes natural Re and Pa to the scale of Malkauns[7].
2. Meend is musical ornament which can be defined as glide from one note to another. Specific notes when used in succession with speed can create a unified and connected effect. This is known as gamak. Andolan is the slow oscillation of a note that starts from a fixed note and touches the periphery of the adjacent note. Murki is a cluster of notes that sounds like a short, subtle taan (a rhythmic pattern of raga notes practised with different permutation). When a knot or cluster of notes is sung or played in order to decorate or embellish another note, it is called khatka[8][9].

2. METHODOLOGY

2.1 Probability-unconditional and conditional

There are two general approaches to estimation that could be used with any probability model, either the observed data likelihood or complete data likelihood. The observed data likelihood is constructed by considering the detection history for each surveyed unit and determining the unconditional probability of observing the data. This is calculated by summing the probabilities for all possible outcomes that could have resulted in the observed detection history at a unit.

A fair die is about to be tossed. The probability that it lands with '5' showing up is 1/6; this is an unconditional probability. But the probability that it lands with '5' showing up, given that it lands with an odd number showing up, is 1/3 (as there are three odd possibilities 1, 3 and 5 out of which one is favourable); this is a conditional probability. In general, conditional probability is probability given some body of evidence or information, probability relativised to a specified set of outcomes, where typically this set does not exhaust all possible outcomes.

Remarks : It can be argued that all probability is ultimately conditional - after all, whenever we model a situation probabilistically, we must initially delimit the set of outcomes that we are prepared to countenance[10].

2.2 Motif and its analysis

A motif is a short melodic thought-more limited than an expression-that happens frequently in a piece of music. A short melodic thought may likewise be known as a motif, a rationale, a cell, or a figure. These little pieces of a song will show up over and over in a piece of music, once in a while precisely the equivalent and in some cases changed. At the point when a motif returns, it very well may be more slow or quicker, or be in an alternate key. It might restore "upside down" (with the notes going up rather than down, for instance), or with the pitches or rhythms adjusted. A harmonic motif is a progression of harmonies characterized in the theoretical, that is, without reference to song or musicality. A melodic motif is a melodic recipe, set up without reference to spans. A musical motif is the term assigning a trademark cadenced recipe, a reflection drawn from the cadenced estimations of a melody[12].

The disclosure of continuous melodic patterns (motifs) is a significant issue in musicology. In music, we can discover a few substances that can be repeated, for example, notes, spans, rhythms, and symphonies movements. All in all, music can be viewed as a line of melodic substances, for example, notes or harmonies on which design acknowledgment procedures can be applied. We can characterize a melodic theme as the littlest significant song component. When in doubt, motifs are gatherings of notes no longer than one measure. In human discourse, a motif is a word. Similarly, as the sentences comprise of words, motifs structure melodic expressions. A song is framed by a few primary motifs, which are repeated, created, and restricted one against another inside the tune development[13].

In order to formalise melody composition, we need to analyse fundamental melodic concepts such as motifs and their relations. This research presents a comparative analysis of musical motifs in *bandishes* (song like raga compositions) of three different ragas of the same *thaat*.

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1 Empirical Unconditional Probability

It is necessary for any raga to have higher probability of occurrence for its important notes and lower probability for its not so important notes given that the probability being calculated is unconditional. The unconditional probabilities of the discussed three ragas from the thaat Bhairavi have been computed. Due to shortage of space the complete notations of the bandishes of respective ragas have not been provided here. The references for the same have been mentioned for the validation if it incites interest to one.

3.1.1 Raga Bilaskhani Todi:

The ascent omits Ma and Ni. Although Pa and Sa are often avoided in the descent, most phrases end on these tones. Ga and Dha are the important tones, and are intoned with a light oscillation. Most movements are in the lower middle octave[14].

The ascent of Bilaskhani Todi is similar to that of raga Bhupal todi , but its descent is different. Although Bilaskhani todi has the same tone material as raga Bhairavi and Asavari, its melodic characteristics are completely different[6].

The *bandish* here taken for analysis is *"Jag Dambika Ambika"*[15]. Total notes used in this bandish are 89 which is our sample space. Probability for each of the notes is given in Table 1.

Table 1. Unconditional probability of notes of raga Bilaskhani Todi

Notations Used	Observed Frequency (O)	Probability = O/89
Sa	21	21 / 89
Re (komal)	16	16 / 89
Ga (komal)	12	12 / 89
Ma	6	6 / 89
Pa	8	8 / 89
Dha (komal)	20	20 / 89
Ni (komal)	6	6 / 89

A moment's reflection on Table 1 suggests that the notes Sa and Dha bear the highest and second highest probabilities respectively. The note Dha being with the second highest probability is in alignment with its theoretical nature. But the experimental results of the notes Ga and Sa do not quite go along with the theoretical results. Theoretically Dha and Ga are vadi and samvadi and hence important tones but experimentally Ga is occurring with only the fourth highest probability. Similarly Sa having the highest probability among all shows that the note having the highest probability need not be the most important note in a raga. In fact, Sa is only the tonic---a reference point from which other notes are realised---while vadi-samvadi are typical of the raga.

3.1.2 Raga Bhairavi:

Raga Bhairavi belongs to Bhairavi thaat. Although anytime is suitable for rendering this raga, it is generally considered a late morning Raga, and traditionally is often the last raga performed at a music session or concert. Shuddh Bhairavi uses all the seven notes in the ascending and descending order, Rishabh, Gandhar, Dhaivat and Nishad being komal (flat) and Madhyam being shuddha (full). The derivative ragas out of this structure are grouped under the broad head of Bhairavi thaat.

Tone material: S r g M P d n (small letter implying a flat note)

Bhairavi with the above tone material is known as Shuddha bhairavi, a raga which is rarely performed today. Commonly however, in Bhairavi both natural Re and flat Re are used, and in thumri, dadra and ghazal performances sharp Ma and natural Dha are used as well. In this case musicians may refer to the raga as Sindhu or Mishra Bhairavi. In instrumental music in addition to the above notes natural Ni may appear[5].

The *bandish* here is *"Ab tori banki lo aniyare"* of raga Bhairavi[16]. This bandish has a total 82 notes. The unconditional probability for each of the notes of the Bhairavi bandish is given in Table 2.

Table 2. Unconditional probability of notes of raga Bhairavi

Notations Used	Observed Frequency (O)	Probability = O/82
Sa	13	13/82
Re (komal)	8	8/82
Ga (komal)	14	14/82
Ma	12	12/82
Pa	14	14/82
Dha (komal)	11	11/82
Ni (komal)	10	10/82

Using a Chi-Square goodness of fit test it can be verified that the observed frequencies can be taken to be uniformly distributed. Hence all the notes seem to be important in this raga *bandish*. The expected frequencies are E=82/7 for each note and the formula for calculating Chi-Square statistic for goodness of fit test is $\Sigma(O\text{-}E)^2/E$, summed over all the seven classes (here the seven notes raga notes) which follows Chi-Square distribution with 7-1 = 6 degrees of freedom. With the given data in table 2, calculated Chi-Square = 2.52 which is less than 12.59 = table Chi Square at 6 degrees of freedom and 5% level of significance and hence we may accept the null hypothesis H0 of uniform distribution of the observed frequencies of the notes in the concerned raga *bandish* (the alternative hypothesis H1 supports non-uniform distribution).

Remarks

(i) Square of a standard normal variate is a Chi Square variate with one degree of freedom. Chi-Square distribution has additivity property, *i.e.,* the sum of independent Chi-Square variates is again a Chi Square variate with degree of freedom as the sum of the degrees of freedom of the individual variates appearing in the sum.

(ii) In Chi-Square goodness of fit test, one degree of freedom is lost due to the linear restriction $\Sigma O=\Sigma E$.

3.1.3 Raga Malkauns

Tone material: S g M d n

The ascent and descent can be direct. The movements in this raga are very slow and gliding. Sa and Ma are important notes and in fact the tanpura is tuned in Ma instead of Pa. During slow movements, flat Ga is held with an oscillation (andol). Dha can be sustained as well.

Bandish for the Raga Malkauns is *"Mukh Mor Mor Muskat"*[17]. Total number of notes in this *bandish* is 84. The unconditional probability for each note is given below (Table 3).

Table 3. Unconditional probability of notes of raga Malkauns

Notations Used	Observed Frequency (O)	Probability = O/84
Sa	28	28/84
Re (not permitted)	0	0/84
Ga (komal)	10	10/84
Ma	21	21/84
Pa (not permitted)	0	0/84
Dha (komal)	12	12/84
Ni (komal)	13	13/84

The unconditional probability of notes of this raga has very confirmatory results. The theoretical importance of the notes Ma and Sa are in complete alignment with its experimental results (Ma is actually vadi and Sa samvadi notes in Malkauns). Clearly the more important notes of this raga are directly proportional to their probabilities.

The notes not permitted (vivadi) theoretically have empirically zero probabilities which is also confirmatory.

The anuvadi notes (an anuvadi note is an important note is a raga but not vadi or samvadi) Ga, Dha and Ni have more or less similar probabilities which is acceptable. Their probabilities are neither high nor low. Thus, they cannot be vadi-samvadi nor they can be alpvadi (unimportant note in a raga).

This raga has its all theoretical results being completely mapped with the experimental results.

Remarks

(i) For raga Bilaskhani Todi, the calculated Chi-Square= 19.32 which is more than 12.59 = table Chi Square at 6 degrees of freedom and 5% level of significance and hence we may accept the alternative hypothesis H1 that the notes are not uniformly distributed in the concerned raga bandish.

(ii) Similarly, for raga Malkauns, the calculated Chi-Square = 13.58 which is more than 9.49 = table Chi Square at 4 degrees of freedom and 5% level of significance and hence we may accept the alternative hypothesis H1 that the notes are not uniformly distributed in the concerned raga bandish.

We next move to compute the conditional probability for the notes of each of the ragas.

3.2 Empirical Conditional Probability

As mentioned earlier, conditional probability is the probability of one event occurring with some relationship to one or more other events. In probability theory, conditional probability is a measure of the probability of an event occurring, given that another event has already occurred.

In case of conditional probability, the sample space for the given condition is not the total number of occurrences of a particular event but the number of occurrences of the event fulfilling the given condition.

Here, computing the conditional probability for each raga involves the probability of occurrence of one particular note being followed by another note. To find the probability of note X to be followed by note Y, we need to find out how many times X has appeared in the note sequence which will be our denominator (sample space) and out of those occurrences of X, how many times X is followed by Y which will be our numerator (favourable outcomes). But if the last note in the sequence is an X, then we have to subtract one from the denominator which will reduce the sample space by unity (because we do not have the information of the next transition for the last note!).

3.2.1 Transition Probability Matrix (TPM):

The probabilities associated with various state changes are called transition probabilities. The (stochastic) process is characterized by a state space, a transition matrix describing the probabilities of particular transitions, and an initial state (or initial distribution) across the state space.

The state transition probability matrix of a Markov chain (in music notes are never independent whereby the unconditional probabilities of notes are different from the conditional probabilities; we are assuming a Markov chain of first order implying the probability of the next note depends on the occurrence of the current note) gives the probabilities of transitioning from one state to another in a single time unit[18].

3.2.2 TPM for ragas Bilaskhani Todi, Bhairavi and Malkauns

TPM for different ragas here will give the information about the conditional probabilities of different notes. It will always be a good comparative study for different ragas of the same thaat. Table 4 gives the TPM for raga Bilaskhani Todi.

Table 4. TPM for Raga Bilaskhani Todi

	Sa	**Re (komal)**	**Ga (komal)**	**Ma**	**Pa**	**Dha (komal)**	**Ni (komal)**
Sa	9/20	7/20	0/20	0/20	0/20	4/20	0/20
Re (komal)	4/16	0/16	8/16	0/20	0/16	0/16	4/16
Ga (komal)	0/12	8/12	0/12	0/12	4/12	0/12	0/12
Ma	0/6	2/6	4/6	0/6	0/6	0/6	0/6
Pa	0/8	0/8	0/8	1/6	1/8	5/8	1/8
Dha (komal)	7/20	0/20	0/20	5/20	2/20	5/20	1/20
Ni (komal)	0/6	0/6	0/6	0/6	0/6	6/6	0/6

In the bandish for this raga, the last note was "Sa" so the total sample space for Sa preceding any particular note will be one less than the total frequency of occurrence of "Sa" in the bandish.

Similarly, for the bandish of raga Bhairavi we have "Re" as the last note. Hence, the total sample space for "Re" preceding another note is one less than the observed frequency of "Re" as can be seen in Table 5.

Table 5. TPM for Raga Bhairavi

	Sa	**Re (komal)**	**Ga (komal)**	**Ma**	**Pa**	**Dha (komal)**	**Ni (komal)**
Sa	2/13	4/13	2/13	0/13	0/13	1/13	4/13
Re (komal)	5/7	0/7	0/7	0/7	1/7	0/7	1/7
Ga (komal)	0/14	3/14	5/14	3/14	3/14	0/14	0/14
Ma	0/12	1/12	6/12	0/12	1/12	2/12	2/12
Pa	0/14	0/14	1/14	8/14	2/14	3/14	0/14
Dha (komal)	0/11	0/11	0/11	1/11	7/11	1/11	2/11
Ni (komal)	5/10	0/10	0/10	0/10	0/10	4/10	1/10

For the bandish of raga Malkauns we have Ma as the last note which implies that the total sample space for Ma preceding any other note will be one less than its observed frequency. Table 6 gives the TPM of raga Malkauns.

Table 6. TPM of raga Malkauns

	Sa	**Re (vivadi)**	**Ga (komal)**	**Ma**	**Pa (vivadi)**	**Dha (komal)**	**Ni (komal)**
Sa	16/28	0/28	1/28	2/28	0/28	3/28	6/28
Re (vivadi)	0	0	0	0	0	0	0
Ga (komal)	3/10	0/10	2/10	5/10	0/10	0/10	0/10
Ma	1/20	0/20	7/20	9/20	0/20	3/20	0/20
Pa (vivadi)	0	0	0	0	0	0	0
Dha (komal)	0/12	0/12	0/12	4/12	0/12	2/12	6/12
Ni (komal)	8/13	0/13	0/13	0/13	0/13	4/13	1/13

Comparing the corresponding cells, it is readily seen that the probability of transition from one particular note to another specific note is dependent on the raga. It is also clear that the unconditional and conditional probabilities of notes are different.

Remark : Probability does not directly explain the decision process of the performer. This decision process is always planned. However, from an analytical or a listener's point of view we can take it as the outcome of a stochastic process. And hence we have taken the notes as random. Our analysis actually proves that although the notes are random, they are not independent but dependent. In probability theory, random variables can be dependent or independent. In music, they are always dependent.

3.3 Comparative Study of *Aaroh* and *Avroh* of the three ragas

This study includes the verification whether or not the *Aaroh-Avroh* patterns of the ragas are being followed in their respective *bandishes* taken. The *Aaroh-Avrohs* of the ragas are given below (Table 7):

Table 7. *Aaroh-Avrohs* of the three ragas

Raga Bilaskhani Todi	Raga Bhairavi	Raga Malkauns
• *Aaroh:* S r g P d S'	• *Aaroh:* S r g M P d n S'	• *Aaroh:* S g M d n S'
• *Avroh:* r' n d M g r S	• *Avroh:* S' n d P M g r S	• *Avroh:* S' n d M g S

Note: S' is the Sa of higher octave

It is observed that

(i) The *aaroh-avroh* pattern is followed in all the three raga bandishes. Hence this can be a good measure for raga identification problem.

(ii) The bandish *"Jag Dambika Ambika"*[2] belonging to raga Bilaskhani Todi interestingly follows both the full *aaroh* and *avroh* of the raga once in its notation sequence unlike the other two ragas.

3.4 Comparative study of the Motifs

While the *bandish "Jag Dambika Ambika"* of raga Bilaskhani Todi has several of its motifs repeated only twice or thrice and do not contribute any statistically significant role in creating any precise melody with those motifs (the significance of a motif in monophonic music can be measured by multiplying the length of the motif by the number of times it is repeated), the motif or sequence *"Dha Sa Sa"* has been used for a maximum of six times and hence has a significance measure 3×6=18 being followed by *"Sa Re Ga"* and *"Sa Sa Re"* each repeated five times with significance 3×5=15. The majority of the melody to the *bandish* of raga Bilaskhani Todi is being created by the following motifs:-

- Dha Sa Sa - 6 (significance measure 18)
- Sa Re Ga - 5 (significance measure 15)
- Sa Sa Re - 5 (significance measure 15)
- Ga Re Sa - 4 (significance measure 12)
- Re Ga Re - 4 (significance measure 12)
- Ga Pa Dha - 3 (significance measure 9)

In raga Bhairavi, the bandish *"Ab tori banki lo aniyare"* the frequency of any motif is comparatively less, as the maximum number of times any motif has been repeated is only thrice. Also there are several motifs used thrice such as :

- Ma Ga Ga - 3 (significance measure 9)
- Ga Pa Ma - 3 (as above)
- Ni Dha Pa - 3 (as above)
- Ga Re Sa - 3 (as above)

etc. So we cannot precisely define the melody with many motifs having similar frequency of occurrences. But here we observe that motifs with 4 and 5 notes together are present thrice and twice. These motifs can be considered as source of generating a particular melody to the bandish. They are:

- Ma Ga Ga Pa - 3 (significance measure 12)
- Ma Ga Ga Pa Ma - 3 (significance measure 15: maximum)
- Sa Ni Sa Ga Ma - 2 (significance measure 10)
- Ma Ga Re Sa Re - 2 (significance measure 10)

In raga Malkauns' bandish *"Mukh Mor Mor Muskat"* the motifs are present with comparatively greater frequency. While there are quite a few motifs being repeated thrice or four times in the sequence, *"Sa Sa Sa"* is the maximum used motif in the bandish *i.e.,* 11 times. But Sa being the tonic this motif cannot be considered as typical of the raga. So we concentrate on the other motifs which are:

- Ma Ga Ma - 3 (significance measure 9)
- Dha Ni Sa - 4 (significance measure 12)
- Ni Sa Sa - 4 (significance measure 9)
- Ma Ma Ga - 3 (significance measure 9)
- Dha Ni Sa Sa - 3 (significance measure 12)
- Ni Dha Ma - 3 (significance measure 9)
- Sa Sa ni - 3 (significance measure 9)

It should be noted here that most of the motifs have been considered if they are appearing thrice or more except in the case of raga Bhairavi where the maximum number of times any motif used is thrice so for a sequence of 4 or more than 4 notations, twice occurrences of the longer motifs have also been taken into consideration.

As we can clearly see that no motifs are common in the respective bandishes of ragas, which shows that the ragas are different though belonging to the same parent thaat, yet an interesting observation is that the notes of several motifs used in the bandishes are more or less same with differing in their arrangement (order). This gives an explanation for the ragas belonging to the same thaat having different melodies. The difference in order is crucial among the notes of the motifs as they give rise to the varying melodic structure nature and rasa of the ragas.

4. CONCLUSION

The probabilistic analysis of the ragas carried out in this research work here shows about how the relationship between the theoretical and experimental values of the notes varies among the different ragas. This was validated with unconditional probabilistic analysis. The conditional probabilistic analysis gives an insight about the probability of the notes being followed by any particular note in the sequence and they too vary from raga to raga even within the same *thaat*.

Of the three *bandishes,* only the bandish in raga Bhairavi has notes whose observed frequencies can be taken to be uniformly distributed, as validated by Chi-Square goodness of fit test.

The comparative analysis of the *aaroh* and *avroh* of all the ragas have revealed that the respective *bandishes* are following the *aaroh-avroh* pattern.

The comparative analysis of motifs in the *bandishes* revealed the difference in the melodic structures of the ragas.

Investigating ragas belonging to different *thaat* but having at least some notes in common is reserved for future work.

Ethical Declaration

The authors hereby declare that this research did not receive any funding and that they do not have any conflict of interest. The work was accomplished by the first author under general guidance of the second author.

5. REFERENCES

[1] W. Kaufmann, 1965. Rasa, raaga-maalaa and performance times in North Indian Raagas. *Ethnomusicology,* **9**, 272-291. doi: 10.2307/850238

[2] N. Bhatkhande Vishnu, 1934. *Hindustani Sangeet Paddhati Kramik Pustak Malika,* Sangeet Karyalaya Hathras.

[3] S. Chakraborty, G. Mazzola, S. Tewari and M. Patra, 2014. *Computational Musicology in Hindustani Music,* Springer International Publishing, Switzerland.

[4] A. Jairazbhoy Nazir, 1995. *The Rags of North Indian Music: Their Structure and Evolution,* South Asia Books

[5] https://autrimncpa.wordpress.com/bhairavi/ ---- accessed 9/11/2020

[6] https://autrimncpa.wordpress.com/bilaskhani-todi/ ---- accessed 9/11/2020

[7] https://autrimncpa.wordpress.com/malkauns/ ---- accessed 11/11/2020

[8] B. Roychaudhuri, 2013. *The Dictionary of Hindustani Classical Music,* Motilal Banarasidass reprint.

[9] H. Hirlekar, 2014. *Nuances of Hindustani Classical Music,* Unicorn Books.

[10] P. Bandyopadhyay (ed), M. Forster (ed), M. Gabbay Dov (ed) and J. Woods (ed), 2011. *Philosophy of Statistics,* Vol. **7**, North Holland.

[11] D. Wilkinson, 1980. *Deadly Quarrels - Lewis F. Richardson and the Statistical Study of War,* University of California Press.

[12] https://courses.lumenlearning.com/musicappreciation_with_theory/chapter/motive/ ---- accessed 15/11/2020

[13] A. Jiménez, M. Solana Miguel, F. Berzal and W. Fajardo, 2011. Mining transposed motifs in music, *Journal of Intelligent Information Systems,* **36**(1), 99-115.

[14] P.M. Thakurdas, 1988. *Rag darshan II*, Sanskar Prakashan.

[15] https://www.youtube.com/watch?v=yjQKUVGcbR0 --- accessed 9/11/2020

[16] https://www.youtube.com/watch?v=QL8Er8LjlR4 ---- accessed 11/11/2020

[17] https://www.youtube.com/watch?v=DNK_uJBdAI0 ---- accessed 12/11/2020

[18] O. Ibe, 2013. *Markov Processes for Stochastic Modeling (Elsevier Insights) 2nd Edition*, Elsevier.

The Journal of Acoustical Society of India : Vol. 48, No. 1-2, 2021 (pp. 56-64)

The perception of features and emotions in the *rāgas* for different patterns of beats, bols and rhythms

Gouri Karambelkar[1*], Junmoni Borgohain[2], Raju Mallick[1], Priyadarshi Patnaik[2,3], Joy Sen[4,5] and Damodar Suar[2]

[1]*Advanced Technology Development Centre, Indian Institute of Technology Kharagpur, Kharagpur-721302*

[2]*Department of Humanities and Social Sciences, Indian Institute of Technology Kharagpur, Kharagpur-721302*

[3]*Rekhi Centre of Excellence for the Science of Happiness, Indian Institute of Technology Kharagpur, Kharagpur-721302*

[4]*Department of Architecture and Regional Planning, Indian Institute of Technology Kharagpur, Kharagpur-721302*

[5]*Ranbir and Chitra Gupta School of Infrastructure Design and Management, Indian Institute of Technology Kharagpur, Kharagpur-721302*

e-mail: gourikarambelkar@gmail.com

[Received: 28-11-2020; Revised: 29-12-2020; Accepted: 29-12-2020]

ABSTRACT

In Hindustani Classical Music, there are multiple features which evoke and modify emotions. Rhythm and tempo are two of the most important among them. An extensive literature review suggests that each feature of a *tāla* - beats, vocables, structure, cyclical beat-structure, metre etc. contributes to the total impact on the perception of music composed in a particular *tāla*. In this empirical study, we wish to examine rhythm *(tāla)* aspect in detail. The survey was conducted on four *rāgas* and four *tālas* - *Khamāja*, *Mānda*, *Mālkaunsa* and *Darbāri Kānadā* using *tālas*, *Teentāla*, *Dādarā*, *Ektāla* and *Choutāla*. Eighty-five participants were instructed to listen to the pair of clips in sequence, focus on the *tāla* of the music and respond. Results showed a clear trend. The *rāgas* in different talas express a gamut of emotions and significantly relate to the selected four features such as complexity-simplicity, slowness-fastness *etc*. The ragas exhibiting talas like Ektala and Teentala significantly model the features of complexity and exciting. On the other hand, *rāgas* in *Choutāla* and *Dādarā* are closely associated with features like slowness and less positivity. Thus, results show that the *tāla,* in itself, is prominent in determining the perception of different features even when the tempo is standardized.

© 2021 Acoustical Society of India

1. INTRODUCTION

Tāla literally means a clap but in Indian music, *tāla* (rhythm) embraces the time dimension of music refers to the musical meter - the means by which musical rhythm and form are guided and expressed[1]. It is an important phenomenon in Hindustani Classical Music (HCM) because it has three main functions: (a) time measurement and time division, (b) accentual pattern / form and (c) cyclicity. According to the traditional Sanskrit saying *'shrutir mātā layah pitā'*, meaning *shruti* (pitch) is the mother and the *laya* (*laya* here is meant in the rather more usual sense of rhythm as regularity and tempo) is the father[2]. Each *tāla* has different number of *mātrās* (beats), different bols (vocables), thus the structure of every *tāla* is unique. Choosing appropriate *tāla* for the compositions in HCM depends basically on the metre of the composition and the nature of the selected *tāla* defines the genre of the composition. In Indian tradition, each *swara* (note) evokes particular emotion[3]. Thus, each *rāga* with a combination of *shuddha* (natural), *komala* (flat) and *teevra* (sharp) *swaras* conveys a dominant emotion with a group of other subsidiary emotions[4]. Along with the *swaras*, as quoted above, all the four aspects of rhythm - rhythm, metre, timing and tempo - also play an important role in perceiving emotions of a music piece[5]. In this study, we wish to look at how the same *rāgas* rendered in different *tāla*s impact the perception of the affective content of the music.

1.1 Rhythm

Rhythm, in music, is the placement of sounds in time. In its most general sense, rhythm (Greek rhythmos, derived from rhein, *"to flow"*) is an ordered alternation of contrasting elements. There are three elements of rhythm (a) beat, (b) tempo and (c) rubatoor, the temporary tightening or slackening of beat[6]. Rhythm covers everything pertaining to the time aspect of music. It includes effects of beats, accents, measures or bars *etc*.[7]. A time signature in Western music notation has two components, specifying between them the number of beats forming each bar and the note value used to notate this beat.

1.2 Metre

Metre is another one of the core concept which is used to describe rhythmic phenomena. Metre is described in two ways by the musicologists - either as a pattern of strong (*sama*, *tāli* and and *bhari* segment of the *tāla* cycle in HCM), weak beats (*khāli* or *kāla* portions of the *tāla* cycle in HCM) or as a grouping of beats (*vibhāgas* of the *tāla* cycle in HCM) for the purpose of measuring time[2]. Although there is a similarity in many concepts of rhythm and *tāla* such as beats *(mātrās)*, grouping the beats in to bars (parallel to the khandas of a *tāla* in HCM), metre *(chhanda) etc*. the main difference in Indian and Western systems is the concept of rhythm is linear in Western system. There are beats, bars and pattern of strong and weak beats are repeated but these beats are not cyclical.

1.3 *Thekā*

Thekā is another unique concept of HCM, which has to be understood while discussing about the structure of a *tāla*. A *thekā* is a basic rhythmic phrase of a particular *tāla*. *Thekā* is an audible indication of a *tāla*. Within the same *tāla*s (total number of *mātrās* remains the same), use of the *thekā* varies greatly. The two *thekās* of the same *tāla* with different / same *khanda*- *vibhāga* (division of beats) with different bols can lead to change in the perception of a *tāla*[2].

The aim of this study is to investigate the perception of stylistic features and emotions in the *rāgas* when they are composed in different patterns of *mātrās* or *tāla*s.

2. METHOD

2.1 Participants

Eighty-five (M=18.7, SD=1.05) undergraduate students studying at the Indian Institute of Technology, Kharagpur, participated. Out of the total participants 92.9% were males and 7.05% were females.

2.2 Design of the interfaces

Similar platforms were used in designing the interfaces as discussed earlier. Here too, the intensity rating interface (Fig. 2) was adopted to assess the emotions perceived in the music clips. Another interface (Fig. 1) was designed to assess the features perceived in the clips.

To describe the clip after listening, choices for selecting an adjective were given in the interface. They were such that the way a novice describes a music piece - exciting or less exciting, slow or fast in laya, simple and complex in its *tāla* etc. This helped us to identify the features of these *tālas* which indicate change in the perception of the laya, complexity and simplicity of *tālas*.

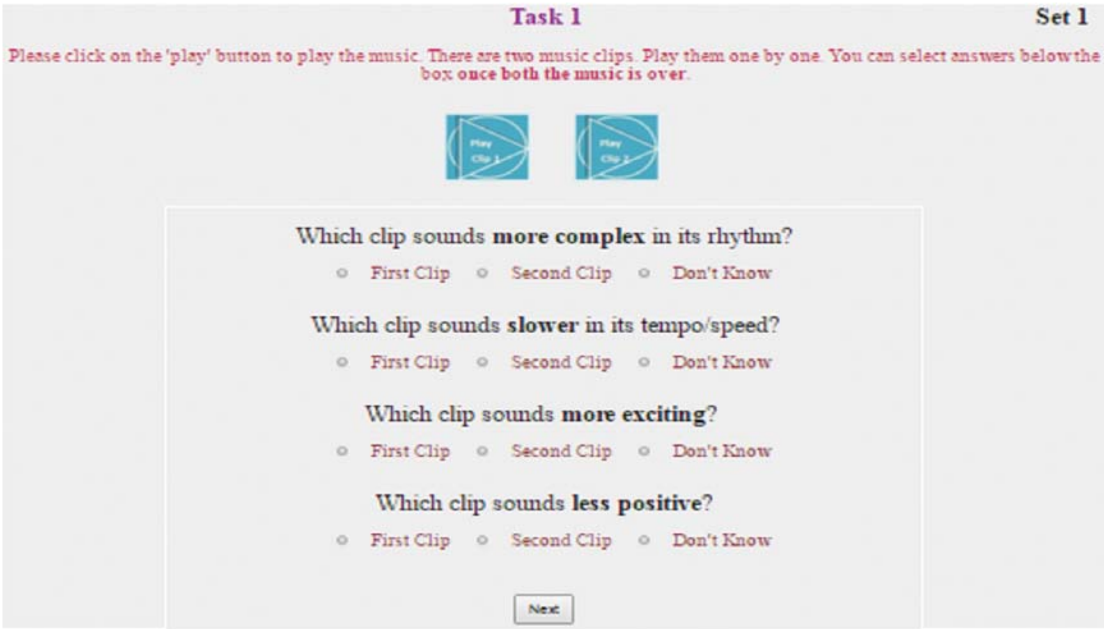


Fig. 1. Interface for identification of features

Fig. 2. Emotion-rating interface

2.3 Stimuli used

For this study, four *rāgas* were selected, namely *Khamāja*, *Mānda*, *Mālkaunsa* and *Darbāri Kānadā*, which were classified on the basis of *tāla* for the study. In pair 1, for comparing the perception of small and large *tālas Dādarā* and *Teentāla* were selected for *rāgas Khamāja and Mānda*, and in pair 2, for studying the perception of *tālas* having same number of *mātrās*, two *thekās* of *tāla Ektāla* - *Ektāla* and *Choutāla*- were selected for *Mālkaunsa* and *Darbāri*. A total of eight clips were selected for the study.

All the compositions were in the form of a *tarānā* (a form consisting of meaningless syllables) to avoid the impact of lyrics on the audience. The compositions were composed keeping the *chalan* of the *rāgas* intact. For accompaniment, electronic *tablā* and *tānapurā* were used. For pair 2, the same composition was rendered in the *tālas Ektāla* and *Choutāla*. The tempo was kept constant at *madhya laya* or medium beat tempo for all the clips.

Table 1. Details of the stimuli used.

	Rāga	***Rāga* details**	**Type of *Rāga***	***Tāla***	**Total Number of *Mātrās***	**Division of *Mātrās***
Pair 1	*Khamāja*	*Komala nishāda* (flat 7th)	*Rāga* used for semi-classical and light genres	*Dādarā*	06	03-03
	Mānda	All *shudhha swaras*	*Rāga* used for semi-classical and light genres	*Dādarā*	06	03-03
	Khamāja	*Komala nishāda* (flat 7th)	*Rāga* used for semi-classical and light genres	*Teentāla*	16	04-04-04-04
	Mānda	All *shudhha swaras*	*Rāga* used for semi-classical and light genres	*Teentāla*	16	04-04-04-04
Pair 2	*Mālkaunsa*	*Komala gandhāra, dhaivata, nishāda* (flat 3rd, 6th, 7th)	*Rāga* used for *Khayāla* genre	*Ektāla*	12	02-02-02-02-02-02
	Darbāri	*komala gandhāra,* dhaivata *nishāda* (flat 3rd, 6th, 7th)	*Rāga* used for *Khayāla* genre	*Ektāla*	12	02-02-02-02-02-02
	Mālkaunsa	*komala gandhāra,* dhaivata *nishāda* (flat 3rd, 6th, 7th)	*Rāga* used for *Dhrupada* genre	*Choutāla*	12	02-02-02-02-02-02
	Darbāri	*komala gandhāra, dhaivata nishāda* (flat 3rd, 6th, 7th)	*Rāga* used *Dhrupada* genre	*Choutāla*	12	02-02-02-02-02-02

2.4 Procedure

As per the objective, to examine the perception of expressive features and emotions when rhythms change in a *rāga*, two major tasks were given in the interface. The first task (Feature rating task) was divided into four sub-tasks (eight clips in the interface). Participants were instructed to listen to a pair of clips (for example- *Darbāri Ektāla* and *Darbāri Choutāla)* one after the other, focus on the *tāla* of the music and on the basis of comparison, answer the four questions that followed each pair of clips keeping in mind variations in rhythm, metre and *thekā* of the *tālas*, which were identified as relevant above. The features that are most affected by these features, based on tradition, were explored through the following questions (see Fig. 1):

a. Which clip sounds more complex in its *tāla*?
b. Which clip sounds slower in its laya (tempo/speed)?

c. Which clip sounds more exciting?
d. Which clip sounds less positive?

Options of choice included first clip, second clip and don't know

After the completion of the first task (rating of features), they were redirected to the CMR task. The users in this task listened to the same music clips as in the earlier task. An instruction page (Fig. 3) at the beginning gave the roadmap to the participants to complete the rating task.

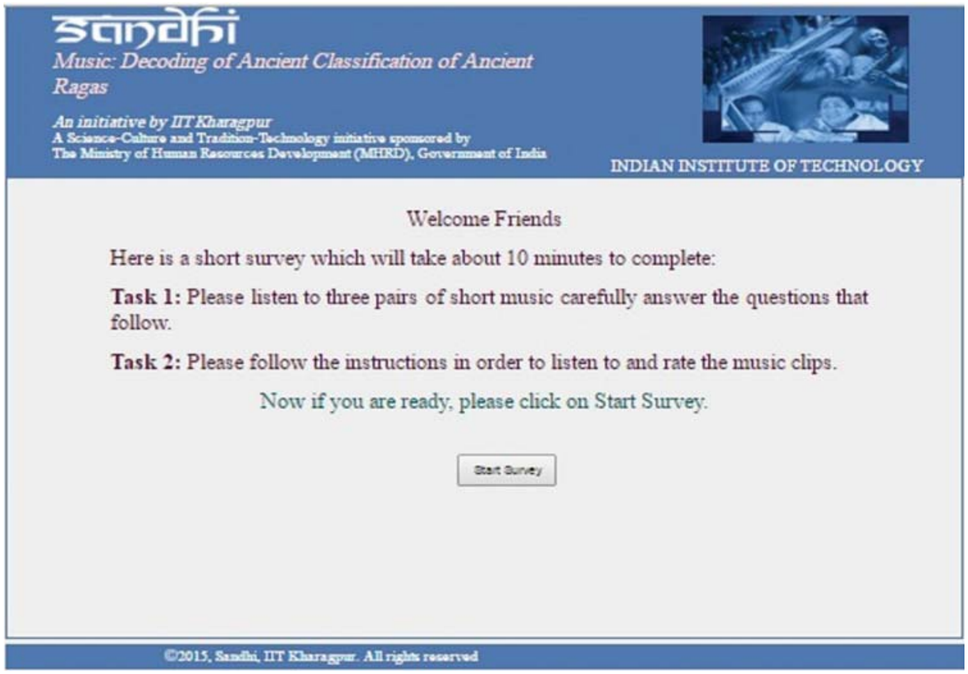


Fig. 3. Instruction page for Experiment

3. DATA ANALYSIS

Chi-square was estimated to find if significant differences existed in choosing the perceptual attributes when rendered in different *tālas*. Discriminant Analysis was utilized to analyse the relation between the emotions and the perceptual attributes.

4. RESULTS AND DISCUSSION

The purpose of this study was to analyse the change in the perception of the different features when the music was rendered in different *tālas*. A related aim was to see whether different emotions expressed correlated with the features. Four specific *tālas* were analysed for four *rāgas*: *Ektāla, Chautāla, Dādarā,* and *Teentāla*. Four features that served as the dependent variables were: Slow clip, complex clip, less positive, and exciting. Participants listened to clips in different rhythms in pairs and answered to the features in the interface. Pearson's Chi-Square test was utilized between pairs of clips to identify significance at a level of .001.

As the feature's percentages indicate (see Table 2), clips in *Ektāla, Darbāri Kānada* were perceived as complex (36.4%) and exciting (50.3%). Similar features were perceived in *Mālkaunsa Ektāla*, where complexity accounted for 35.8% and exciting for 49.7%. While the *rāga* was rendered in *Chautāla*, it was identified as slow (44.2%) and less positive (36.1%) in *rāga Darbāri* and *Mālkaunsa* also possessed the same attributes; 43.8% identified it as slow and 35% of them identified it as less positive.

Table 2. Chi-square Test and Descriptive Statistics for Features in Different Rhythms

Clips	Rhythm	Features				Pearson Chi-Square
		Slow	Less Positive	Complex	Exciting	
Darbāri	*Ektāla*	7 4%	16 9.2%	63 36.4%	87 50.3%	183.053***
Darbāri	*Choutāla*	92 44.2%	75 36.1%	32 15.4%	9 4.3%	
Mālkaunsa	*Ektāla*	6 3.6%	18 10.9%	59 35.8%	82 49.7%	165.615***
Mālkaunsa	*Choutāla*	89 43.8%	71 35%	33 16.3%	10 4.9%	
Khamāja	*Dādarā*	54 31.2%	59 34.1%	41 23.7%	19 11%	58.931***
Khamāja	*Teentāla*	23 14.4%	21 13.1%	47 29.4%	69 43.1%	
Mānda	*Dādarā*	46 30.9%	42 28.2%	26 17.4%	35 23.5%	16.958***
Mānda	*Teentāla*	28 16.6%	36 21.3%	56 33.1%	49 29%	

Note: ***p <.001

For the *rāgas* in *Teentāla*, the perceived the features were exciting and complex. For instance, in *Khamāja* the percentage response to perceived complexity was 29.4%, and exciting was 43.1%. In the same way, ratings on complexity was 33.1% and exciting was 29% in *Mānda*. Contradictorily, clips in *Dādarā* rhythm such as *Mānda* and *Khamāja* were observed as less positive and slow.

Among the clips for *Ektāla* and *Choutāla*, *rāga Darbāri* was perceived sad (33.7%) and calm (28.88%) for *Choutāla* and exciting (20.50%) and calm (18.92%) whereas for the same *tālas*, in the other clips of *rāga Mālkaunsa*, *Ektāla* clip was perceived as sad (20.64%) and exciting (11.52%) and *Choutāla* as calm (24.43%) and exciting (21.71%). For the *tālas Dādarā* and *Teentāla*, *rāga Khamāja* was rated romantic (40.65%) and calm (18.18%) and happy (20.55%) and exciting (15.55%) respectively. *Rāga Mānda* was perceived romantic (26.38%), exciting (18.29%) and romantic (52.98%), happy (14.57%) for the *tālas Dādarā* and *Teentāla* respectively (see Table 3).

Table 3. Emotions perceived in the *rāgas* (values in %)

	Darbāri Ektāla	*Darbāri Choutāla*	*Khamāja Dādarā*	*Khamāja Teentāla*	*Mālkaunsa Ektāla*	*Mālkaunsa Choutāla*	*Mānda Dādar?*	*Mānda Teentāla*
Calm	18.92	28.88	18.18	12.22	29	24.43	16.17	12.56
Anger	3.78	7.77	3.53	5.83	7.77	5.88	1.27	2.01
Exciting	20.50	2.22	7.07	15.55	11.52	21.71	18.29	9.54
Wonder	8.51	4.81	5.8	10.27	5.89	12.66	8.08	4.27
Sad	14.19	33.7	8.08	2.77	20.64	11.31	7.65	3.51
Fear	5.0	7.03	3.28	2.77	10.45	6.78	1.27	1.2
Happy	14.19	4.07	10.85	20.55	4.55	4.97	15.74	14.57
Romantic	8.51	3.7	40.65	25	4.28	3.61	26.38	52.98
Don't know	2.83	3.7	0.99	1.94	2.41	4.07	3.4	1.5
Other emotions	3.15	4.07	0.66	3.05	3.21	4.52	1.7	1.75

Discriminant analysis was performed to determine whether the features assigned to four groups namely, complex, exciting, less positive and slow clip could be classified and categorized by the participants on the basis of the emotions. Focusing on the results of the chi-square test, discriminant analysis was done only for those features that had significant ratings in the *rāga*. Two classification functions (Yes / No) were used to assign cases into each group, and two classification scores were computed for each function. The standardized canonical discriminant function coefficients of the predictor variables were taken into account. It measures the relative importance of the selected variables (*i.e.,* the larger absolute value of the coefficient corresponds to greater discriminating ability).

Discriminant analysis indicated that when the feature of complexity was perceived various emotions were associated with it (Table 4). In the *rāga Darbāri Ektāla* wonder (0.856) was the most powerful discriminating variable, followed by fear (0.571) for the complexity perceived. *Mālkaunsa Ektāla* was

Table 4. Discriminant analysis of Emotions and Features.

	Significant features	**Significant Predictors (Emotion)**	**Standardized Canonical Discriminant Function Coefficients**	**Canonical correlation**
Darbāri Ektāla	Complex	Wonder	0.856	0.359
		Fear	0.571	
	Exciting	Fear	0.792	0.437
		Exciting	0.268	
Mālkaunsa Ektāla	Complex	Fear	0.243	0.315
		Anger	0.128	
	Exciting	Happy	0.699	0.208
		Wonder	0.408	
		Romantic	0.404	
Mālkaunsa Choutāla	Slow	Romantic	0.790	0.163
		Happy	0.578	
		Calm	0.531	
	Less positive	Happy	0.629	0.334
		Exciting	0.575	
Khamāja Dādarā	Slow	Happy	0.572	0.364
		Romantic	0.308	
	Less positive	Calm	0.198	0.283
		Wonder	0.184	
Khamāja Teentāla	Complex	Anger	0.609	0.286
		Romantic	0.379	
		Exciting	0.294	
	Exciting	Romantic	0.661	0.354
		Fear	0.607	
Mānda Dādarā	Slow	Wonder	0.521	0.296
		Happy	0.457	
	Less positive	Anger	0.500	0.264
		Happy	0.414	
Mānda Teentāla	Complex	Anger	0.669	0.276
		Happy	0.539	
		Exciting	0.509	
	Exciting	Sad	0.799	0.431
		Romantic	0.697	
		Calm	0.516	

affiliated with fear (0.243) and anger (0.128). Similarly, *Khamāja Teentāla* was linked with emotions such as anger (0.609), romantic (0.379) and exciting (0.294). *Mānda Teentāla*, on the other hand, had anger (0.669), happy (0.539) and exciting (0.509) as dominant emotions responding to the inherent characteristic in the clip.

Another feature recognized by participants was exciting. The *rāga Darbāri Ektāla* indicated emotions like fear (0.792), and exciting (0.268) predicted this feature, whereas, *Mālkaunsa Ektāla* demonstrated associations of this feature with happy (0.699), wonder (0.408) and romantic (0.404). Romantic (0.661), and fear (0.607) were the most powerful indicators in *Khamāja Teentāla*. However, for *Mānda Teentāla* sad (0.799), romantic (0.697), and calm (0.516) were the most significant predictors for the exciting clip which makes it a bit contradictory.

Discriminating variables like romantic (0.790), happy (0.598) and calm (0.531) were included in the slow-clip feature for *Mālkaunsa Choutāla*. Happy (0.572) and romantic (0.308) were significant predictors in *Khamāja Dādarā*. Lastly, *Mānda Dādarā* showed wonder (0.521) and happy (0.457) as the significant discriminating variables.

For the less positive clip feature happy (0.629) and exciting (0.575) were the significant discriminating variables in *Mālkaunsa Choutāla*. Discriminant analysis of the less positive feature revealed calm (0.198), romantic (0.194) were the significant predictors in *Khamāja Dādarā*. Finally, for *Mānda Dādarā* Anger (0.500) and Happy (0.414) were the most associated emotions. Significance was not achieved in *Darbāri Choutāla* and hence not reported in the analysis.

It was suggested by Wieczorkowska *et al.*[4] that a *rāga* comprises many phrases and motives and evokes a variety of emotions. Our interfaces have been able to capture this significant aspect of the *rāga*. The results show close connection of different emotions with the features as well as the *rāgas*. Mathur *et al.*[8] stated that all *rāgas* universally created a calming effect and anger ranked as the lowest emotion. This notion is challenged in our study. It was observed that *rāgas* exhibit a gamut of emotions and the emotion anger was reflected where the feature of complexity was perceived in the *rāga*.

5. CONCLUSION

This study explored the role of *tāla* in music perception. Experimental research has uncovered several musical features that represent emotions, but the variability of rhythm in the same *rāga*- particularly whether changes in emotional responses occur, motivated this study. While further exploration is required with more tempo variations and *rāga* examples, preliminary findings show a distinctive impact of rhythm of both emotions and expressive features. Moreover, perceived valence of emotions also correlated with expressive features. If a clip is rated as positive, it correlates with excitement and perceived complexity. On the other hand, negatively rated clips lead to perception of less complexity and slowness. Interestingly, while primary emotions seem not be influenced by the rhythm, the associated secondary emotions - often called *sanchāribhāvas* in Indian tradition - show significant change, indicating what apart from tempo, rhythm does impact perception of emotions in HCM. Future research may need to address issues of the way rhythm features change with tempo across a wider range of *tālas* and *ragas*.

6. REFERENCES

[1] *Tāla,* 2020. In Wikipedia. https://en.wikipedia.org/wiki/Tāla _(music)

[2] M. Clayton, 2000. *Time in Indian music: rhythm, metre, and form in North Indian rag performance.* Oxford University Press.

[3] K. Brihaspati, 1991. *Bharat ka Sangeet Siddhant*. (2nd Ed.). Lucknow: Uttar Pradesh Hindi Sansthan.

[4] A.A. Wieczorkowska, A.K. Datta, R. Sengupta, N. Dey and B. Mukherjee, 2010. In search for emotion in Hindusthani vocal music. *In Advances in music information retrieval.* Springer Berlin Heidelber, pp. 285-304.

[5] H. Honing, 2013. *Structure and interpretation of rhythm in music.* The Psychology of Music. Elsevier Inc. pp. 369-404.

[6] Rhythm. 2020. *In Encyclopædia Britannica online*. Retrieved from https://www.britannica.com/art/rhythm-music.

[7] P. Scholes, 1996. Rhythm. In The Concise Oxford Dictionary of Music. (pp. 482-483).

[8] A. Mathur, S.H. Vijayakumar, B. Chakrabarti and N.C. Singh, 2015. Emotional responses to Hindustani raga music: the role of musical structure. *Frontiers in psychology*, **6**, 513.

The Journal of Acoustical Society of India : Vol. 48, No. 1-2, 2021 (pp. 65-67)

Madhyama rahitha "Melakartha" scales in carnatic music

Rama Bhat
Mechanical, Industrial and Aerospace Engineering,
Concordia University, Montreal, Canada
e-mail: rbhat2007@gmail.com

[Received: 05-10-2020; Revised: 10-12-2020; Accepted: 15-12-2020]

ABSTRACT

The Carnatic classical music system which is prevalent in South India uses 12 semitones within an octave. Principal melodies are denoted by seven syllables in the solfa system named S, R, G, M, P, D, and N (pronounced Sa, Ri, Ga, Ma, Pa, Da, and Ni, respectively). The first note S, and the perfect fifth, P, are fixed in frequency in a principal melody scale according to convention. In any principal melody scale (Janaka Raga), either the perfect fourth, M1, or the augmented fourth, M2, will be present exclusively. The second, third, sixth and the seventh notes can have three variations each, some of which overlap as follows: R1, R2 = G1, R3 = G2, G3, D1, D2 = N1, D3 = N2, N3. With the above notation and using each solfa symbol S, R, G, M, P, D, N only once in the ascending and descending order, traditionally 72 principal melody scales have been formed. Each one of them can have many derived scales by excluding some notes, or using the notes in a convoluted fashion. By including both M1 and M2, 36 new principal melody scales have been proposed recently, named Dvimadhyama ragas. The notes R, G, D, and N have three variations each. Similar to the idea of the two M notes simultaneously occurring in a principal scale, if two notes with solfa symbols, R, G, D, and N occur together, they will form 24+24=48 additional principal scales. In order to avoid more than 7 distinct notes occurring in each principal scale, the notes M1 and M2 are completely removed. Since these principal scales are obtained by removing away the M1 and M2 notes, they are named "Madhyama Rahitha" scales. Shruthi can be held with S and P notes as usual.

1. INTRODUCTION

The principal melody scales in Carnatic music system were first propounded by Ramamatya in his work Svaramelakalanidhi in 1550[1]. Venkatamakhi proposed in the 17th century in his work Caturdandi Prakaasikaa a new set of principal melody scales known as melakartha scales[2, 3]. There were some bold claims made by defining 12 semitones in an octave at that time to arrive at 72 principal melody scales. The double counting of R, G, D and N etc. and his exclusive selection of either the M1 or the M2 notes in a scale are arbitrary. However, today these 72 principal melody scales have gained acceptance. The 12 notes in an octave with their corresponding frequencies are given in Table 1. The lowest note is held at 500 Hz. The overlapping notes are also indicated in Table 1.

© 2021 Acoustical Society of India

Table 1. Overlapping Notes in an Octave and Corresponding Frequencies

Notes	Frequency in Hz
S*	1000.00
N3	944.94
N2, D3	891.91
N1, D2	841.85
D1	794.60
P	750.00
M2	707.10
M1	667.42
G3	630.00
G2, R3	594.60
G1, R2	561.23
R1	529.73
S	500.00

1.1 Vikrutha Panchama Melakartha Scales

By including both M1 and M2, 36 new principal melody scales have been proposed recently, named Dvimadhyama ragas[4]. More recently, P has been completely eliminated in the new principal scales in order to keep the number of notes to 7 so as to limit the maximum number of distinct notes to 7 in a principal scale[5, 6]. The original note P is discarded and the unused solfa syllable P is attached to the M2, in order to keep the total number of notes to seven. They are named as Vikrutha Panchama scales. By varying the solfa syllables R, G, D, N as before will result in 36 new principal melody scales. Including Venkatamachi's 72, this will bring the total fundamental melody scales to 108.

The note P is normally taken as fixed note. But this has not been always like that. In order to use both M1 (5^{th} semitone) and M2 (6^{th} semitone) in a new set of scales, the note P (7^{th} semitone) is abandoned and the 7th semitone is named as P, the frequency corresponding to M2. This way all the seven syllables, S, R, G, M, P, D, and N are used in the set. Since the perfect fifth note is discarded, it cannot be used in the shruthi, and the perfect fourth should replace the perfect fifth in the shruthi. This scheme will generate 36 more scales. The first six of these are listed in Table 2. The remaining 30 scales are obtained by changing the R, G, D and N notes appropriately.

Table 2. The first Six Vikrutha Panchama Melakartha Scales (Numbers refer to the semitone number)

S	R1	G1	M1	M2	D1	N1
S	R1	G1	M1	M2	D1	N2
S	R1	G1	M1	M2	D1	N3
S	R1	G1	M1	M2	D2	N2
S	R1	G1	M1	M2	D2	N3
S	R1	G1	M1	M2	D3	N3

The total number of scales will now be 108 including the 72 propounded by Venkatamakhi. In Hinduism the number 108 has a special significance.

1.2 Proposed Additional Melakartha Ragas Using Madhyama Rahitha Scales

An additional 48 melakartha scales can be created when we sequentially keep two of the R, G, D, and N notes repeated. Thus the scheme in Table 3a will provide 4×6=24 scales and Table 2b) will provide 24 scales. For example, from Table 3a, the first of these scales is: S, R1, R2, G2, P, D1, N1.

Table 3. Madhyama Rahitha Scheme

(a) Repeated R and Variations in D and N		
	R1, R2, G2	D1, N1
	R1, R2, G3	D1, N2
	R1, R3, G3	D1, N3
	R2, R3, G3	D2, N2
		D2, N3
		D3, N3
(b) Repeated N and Variations in R and G		
	D1, N1, N2	R1, G1
	D1, N1, N3	R1, G2
	D1, N2, N3	R1, G3
	D2, N2, N3	R2, G2
		R2, G3
		R3, G3

2. CONCLUSION

Using S and P fixed and choosing either M1 or M2, Venkatamakhi proposed 72 principal ragas called melakartha ragas. Using both M1 and M2 in a scale A, new set of 36 melakartha scales, named Dvimadhyama scales, were proposed. They could not be strictly considered a melakartha scheme because they consisted of 8 notes in an octave. Subsequently by removing P from the scales the 36 ragas could be considered melakartha ragas because the scales consisted of 7 independent notes in the octave. This added 36 new melakartha scales. When the shruthi is held with S, M2 and S* (S in higher octave), the scales are named as vikrutha panchama scales. The addition of these 36 scales makes the total number of melakartha scales to be 108. The numbers of derived scales from these additional 36 scales are unlimited.

When two overlapping neighboring notes occur in a scale, there are corresponding 48 melakartha scales. Altogether the Venkatamakhi scales, vikrutha panchama scales and the 48 madhyama rahitha scales result in 156 melakartha scales. Some of the derived (Janya) ragas (consisting of less than 7 independent notes taken from any melakartha scale) are common for more than one melakartha scale and result in innumerable raga system.

3. ACKNOWLEDGEMENT

Discussions with Vidwan Vittal Ramamurthy, Chennai, on Dvimadhyama ragas are gratefully acknowledged.

4. REFERENCES

[1] Ramaswami Aiyar and Ramamatya's Svaramelakalanidhi, 1932. The Annamalai University Publications, Chidambaram, India.

[2] R. Sathyanarayana, 2002. Chaturdandiprakasika by Venkatamakhi, Indira Gandhi National Centre for the Arts in association with Motilal Banarsidass Publishers, Delhi.

[3] P. Sambamurthy, 1983. South Indian Music, The Indian Music Publishing House, Madras.

[4] S. Kalyanaraman, 1930-1994.

[5] Rama Bhat, 2007. "Vikrutha Panchama Scales in Carnatic Music Containing both Fourth and Diminished Fifth", Canadian Acoustical Association Conference, Concordia University, Montreal.

[6] https://en.wikipedia.org/wiki/Dwi-Madhyama_Panchama_Varja_Ragas

The Journal of Acoustical Society of India : Vol. 48, No. 1-2, 2021 (pp. 68-75)

Identifying style of vocalist using silence in Hindustani music

Anirban Patranabis, Kaushik Banerjee, Ranjan Sengupta and Dipak Ghosh
Sir C.V. Raman Centre for Physics and Music, Jadavpur University, India
e-mail: anir.thikana@gmail.com

[Received: 19-11-2020; Revised: 14-12-2020; Accepted: 14-12-2020]

ABSTRACT

In Hindustani Music (HM), pause or silence evoke the essence in a raga that controls the finest emotional behaviour. Proper utilisation of musicological ornamentations along with properly and appropriately used silence in between them is the outcome from a perfect rendition of any raga that touches the hearts of music lovers. In a raga rendition properly utilized pause in between phrases evokes an essence of the raga. In HM, a musician has full freedom to improvise (modification in phrases during playing) their composition effectively. So to improvise, a good vocalist must know how to use silence effectively. In this work we have make an attempt to find out the importance of pause in HM in identifying style of an artist. Here we have measured the silence / pause parts of eighty-six musical signals sung by twenty-two eminent vocalists each of them sung same four ragas. An attempt was made to understand their singing style in the light of pauses within their performances. Various note sequences (phrase) and the pause parts were located in each of the sound signals. Total number of pauses and time duration of both pause and phrases were measured. Sequences of phrase arrangement at the onset and offset of each pause were measured. Compiling this information, the behaviour of pauses present in musical signals was evaluated. This analysis can signify certain information about the raga movement and the style of presenting a raga by an artist. This new approach to understand and measure a particular artistic phenomenon is non-trivial to the musicians.

1. INTRODUCTION

Shabda Bramha or *Nada Bramha*,[1] which means the universe had been formed with sound. In other words: transcendental sound. In Hindustani music 'Omkar dhwani' is denoted as 'nada' and music often described as 'Nada-vidya'[2]. Musically, if we pronounce that would be Pa-Sa (fifth note-first note) or Ma-Sa (forth note-first note) with *meend* (uninterrupted) in general. But if there is no silence how can we realize the beauty of sound? And to understand a sound properly whether it is noise or speech or music, in-between silence is necessary. Pause or silence is not only necessary for our breathing but it also creates different meaning in our speech or music. The application of accent or intonation in both speech and music makes it colourful. Pause within music isn't just the canvas upon which music is painted. It's one of the colours on the composer's palette.

What sets Hindustani (North Indian Classical) music apart from other music forms is that each musician (through improvisation) is also a composer and needs to know how to use pause effectively.

© 2021 Acoustical Society of India

We have often thought that a master class should be taught on the role of pause in music because measures of pause are not waiting periods. Therefore, the length of pause, periodicity of occurrence of pause, and the magnitude of pause all might make an effect on the listener. Otherwise it would be monotonous. On the contrary, stopping (pause) for a moment or taking jiffy rest in between two lines or words or *taans* (pattern of musical notes) *etc.* is very noteworthy for vocal, especially classical vocal music. Normally, these jiffy rest or pause occurred either for breath taking or as a surprising element (approach) in melodic or rhythm pattern by the singers. In Hindustani classical vocal music, the amount of pauses may be different from one *gharana* (musical patterns for a definite region) to another; even it might differ (a little bit) from artist to artist as per their breathing power and creative emotional perspective. There are many melodic exercises to increase the power of breathing *(Sans Bhartist Ha)* in all vocal *gharanas* and other than common approaches there are some special kinds of exercises which depict the *gharana's* perspective towards a particular group of ragas.

Hindustani classical music is based on raga and *tala*. To form a raga there are some rules, of which *nayas swara* (resting note), *chalan* (notes movement of *raga* in ascending and descending order), *pakad* (main phrase of a *raga*) and a number of do's and don'ts[3]. There are also some ornaments known as *alankar* like - *meend, gamak, andolan,* variety of *tan etc*. Artists at the time rendition follows all the rules but still their styles could be recognized by the expert listeners. Here style denoted by both individuality and Gharana or school from which he or she belongs. Application of *alankars* differs from one raga to another and from one Gharana to another. That means, in some Gharanas *meend* is the vital *alankar* (Gwalior, Agra Gharana) whereas; some prefers gamak (Kirana, Vindibazar), swift tans (Patiala) *etc*.[4]. But that does not mean that they use only those *alankars;* along with other *alankars* only precedence on those. After a long *talim* (training) a reputed artist usually at the time of rendition always maintains the Gharana's major rules by which they can express the emotion of raga. But at the same time from their experience, imagination, extraordinary talent and inner feeling added more musical ingredients (maybe *alankar*) which may not use in that Gharana or not even in that particular musical genre. And in this way some musicians create their own identity which we recognize as 'style'.

Thus, analyzing those abovementioned *alankars* and other musical ingredients we can identify the style of an artist and his / her Gharana as well. Here we applied the silence and note relation. From the perspective of silence and note relation we can say that each *alankar* has its own pattern (through algorithmic analysis). In Hindustani music, usually, at the time of rendition artist performs extempore where artist's mood plays a vital role. As a result, same raga presentation in different concerts becomes dissimilar to the listeners; so they can enjoy. Generally, this dissimilarity happens by means of use of notes and its duration, nuances of *alankar* applications etc. On the contrary; apart from timbre, approach of notes application along with its *sruti* (microtones) *proyog* (application) are the major characteristics to understand an Artist. Our approach is to understand what 'style' actually means, through the perception of silence and note relations. It is a unique parameter by which we can understand the actual physical aspects of style.

In this study the rests or pauses in between phrases of a musical signal are known as 'pause / silence'. By analyzing these pauses along with the notes in the signal and their duration we can get a picture of the style of a particular artist. Both the sequence of notes and pause together forms the ultimate musical structure. Also the note sequence at the onset of pauses is also an important cue for style analysis of an artist.

2. DATA ACQUISITION

Here in this study, we have recorded the vocal musical pieces sung by twenty-two Hindustani vocalists. Out of these artists, some are very famous, well known and eminent singers of HM. Some have more than fifteen years of experience in HM. Among them, only four are female artists and rest eighteen are male. All recordings were done in a professional recording studio without any interference of sound of any other musical instruments or background music except tanpura drone. We have considered four ragas, viz. Bhairav, Darbari Kannada, Mian-ki-Malhar and Todi. Each of them has sung all these four ragas. Only the aaalap (introductory part of a raga which reflects the complete essence of the raga without any accompanying instruments) part of the said ragas were considered here in this study. Duration of the signal varies from 3 seconds to 6 seconds. All signals were digitised @44100 samples / second (16 bit / sample) using standard software WAVESURFER.

3. METHODOLOGY

Figure 1 shows the spectral view of a silence in between phrases. Non-musical part or void in between two phrases is considered as silence or pause. Such silences were identified and hence their durations and total number of such voids were measured. Some silences are long durational while some are short durational. Pitches were identified and note sequences were measured computationally based on algorithms by A. K. Dutta *et. al.*[5,6,7]. The note sequences and hence the phrase patterns were also identified. Time duration of each silence and each phrases were measured. Average silence durations in second, silence to note ratio[8] and frequency of occurrences of silence were measured. These results were compared among each other and tried to find the uniqueness. At last k nearest neighbour method was applied to find the similarity. Correlation coefficients among time duration of silence and time duration of pauses were also measured.

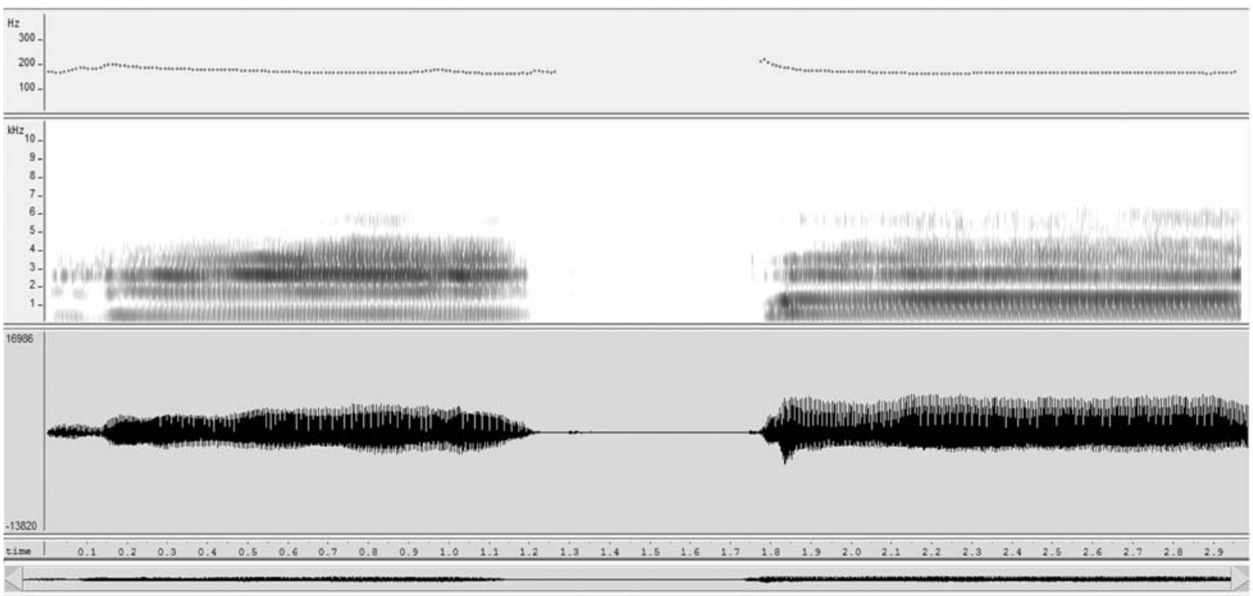


Fig. 1. Pitch, Spectrogram and waveform view of a silence in between phrases

4. EXPERIMENTS AND RESULTS

Table 1 shows the average silence duration and ratio of silence duration to note duration (called as silence to note ratio) for four ragas sung by twenty-two artists. Average silence duration is found to be uniform in the voice signals of the artist ABD and AJO. Silence durations are rarely found to be greater than one second for both the artists in all the four raga renderings. Average silence duration for the artist ABD is 0.99 second and that for AJO is 0.92 second. Silence to note ratio is also found to be uniform in all the sound signals of both the artists. Average silence to note ratio for the artist ABD is 0.11 and that of AJO is 0.1. Such uniformity in silence duration and silence to note ratio are the style stamp of these two artists. Majority of pause is followed by long durational meend (transition between notes). Such pause pattern is an important style stamp for these two artists.

Table 1. Average silence duration in second and silence to note ratio of four ragas of twenty-two artists

	Average silence duration (sec)				silence to note ratio			
	Bhairav	Darbari kannada	Mian ki Malhar	Todi	Bhairav	Darbari kannada	Mian ki Malhar	Todi
ABD	0.957	0.959	0.944	0.898	0.131	0.106	0.104	0.109
AJO	0.955	0.856	0.944	0.825	0.105	0.084	0.115	0.097
AMI	1.603	1.605	1.574	1.743	0.191	0.191	0.187	0.201
ANI	0.456	0.515	1.038	0.796	0.048	0.054	0.116	0.087
ARN	1.592	1.303	1.580	1.518	0.188	0.150	0.188	0.179
ARS	1.321	1.422	1.366	1.373	0.152	0.164	0.145	0.146
ARU	1.222	1.294	1.721	1.786	0.139	0.149	0.208	0.217
ATK	2.148	2.330	2.165	2.263	0.274	0.357	0.327	0.292
DOL	1.227		1.078	1.227	0.140		0.128	0.140
FAL	1.249	1.210	1.289	1.188	0.143	0.132	0.148	0.135
MAS	1.725	1.611	1.679	1.789	0.208	0.192	0.197	0.218
NIR	1.123	1.211	1.147	1.106	0.127	0.151	0.130	0.124
ONK	1.363	0.989	1.864	2.192	0.158	0.110	0.229	0.281
RUP	3.709		1.263	3.246	0.589		0.145	0.481
SAB	2.641	1.155	1.359	1.633	0.359	0.131	0.157	0.195
SAF	1.816	1.731	1.828	2.093	0.222	0.239	0.224	0.241
SAN	1.304	1.788	1.897	1.368	0.150	0.218	0.234	0.158
SUC	0.660	0.857	1.099	1.240	0.071	0.094	0.123	0.142
SUN	2.172	2.258	2.094	2.171	0.266	0.292	0.265	0.277
ULH	1.738	1.243	1.204	0.252	0.210	0.142	0.137	0.026
WAS	1.108	1.183	1.243	1.172	1.100	0.134	0.142	0.133
ZAI	1.277	0.912	1.206	1.016	0.146	0.100	0.137	0.113

Similar average silence duration is observed in all the raga renderings of the artist DOL, FAL, NIR, ARS and WAS. Most of the silence duration is more than one second for the artist DOL, FAL and NIR while no such trend is observed in the renderings of the artist ARS and WAS. Mixed silence duration is observed in the renderings of these two artists. Average silence duration for these four artists are 1.06, 1.23, 1.15, 1.32 and 1.18 seconds respectively. Silence to note ratio is also found to be similar in all the raga renderings of these artists. Average silence to note ratio for the artist DOL, FAL, NIR, ARS and WAS is 0.1, 0.14, 0.13, 0.15 and 0.38. Such uniformity in silence duration and silence to note ratio are the style characteristics of the artists DOL, FAL and NIR. No definite trend in pause pattern is observed in these artists. Pause are applied precedence to or followed by all types of ornamentations like meend, andolan, murki, gamak and khatka.

Uniformity in the average silence duration is observed in all the four ragas of AMI, ARN and MAS. Observing the signals of the artist AMI, it is found that the majority of silence durations are found to be around one second in the raga darbari kannada, mia ki malhar and todi, while it is more than two seconds in the raga bhairav. Majority of silence durations are found to be around one second in all the signals of the artist MAS while no such trend is observed in the signals of the artist ARN. Uniform silence to note ratio is measured in all the four renderings of the artists AMI and MAS. So similar pause duration and silence to note ratio are style parameters of the artist MAS. Silence to note ratio is found to be uniform in the ragas bhairav, mia ki malhar and todi while it is little less in the raga darbari kannada of the artist

ARN. Average silence duration of these three artists are 1.63, 1.5 and 1.7 seconds, while average silence to note ratio is 0.19, 0.18 and 0.2 respectably. Pauses are associated with all types of ornamentations like meend, andolan, murki, gamak and khatka. No definite pattern is observed for these artists.

Similarity in average silence duration and silence to note ratio is observed in the raga renderings of the artists ATK and SUN. Average silence duration is found to be quite high compared to all other artists. Average silence duration of these two artists are 2.23 and 2.17 seconds while average silence to note ratio is 0.31 and 0.27 respectably. Long durational silence followed by or preceded to ornamental notes is definitely a style stamp of the artists ATK and SUN. Majority of pause is found to be after or before the ornaments like meend, andolan and gamak in all the raga renderings.

For the artist ARU, uniformity in both the average silence duration and silence to note ratio are observed for the raga bhairav and darbari kannada with majority of silence durations are less than one second. Also these two parameters are found to be uniform in mia ki malhar and todi for this artist with majority of silence durations are more than one second. For the artists RUP and SAN, uniformity in both the average silence duration and silence to note ratio are observed for the raga bhairav and todi. Also these two parameters are found to be uniform in mia ki malhar and darbari kannada for these two artists. Majority of silence durations are observed to be more than two seconds for the artist SAN while no such trend is observed in the sound signals of RUP. For the artist ZAI, uniformity in both the average silence duration and silence to note ratio are observed for the raga bhairav and mia ki malhar with majority of silence durations are more than one second. Also these two parameters are found to be uniform in darbari kannada and todi for this artist with majority of silence durations are less than one second.

Similarity in average duration of silence are observed in bhairav, darbari kannada and mia ki malhar only but silence to note ratio is found to be similar in all the four raga renderings of the artist SAF. Majority of silence duration of this artist is found to be more than one second. Both the average duration of silence and silence to note ratio are found to be similar in the raga renderings of darbari kannada and mia ki malhar only for the artist ULH. Majority of silence duration of this artist is found to be less than one second. No similarity in silence duration is observed in the raga renderings of the artists ANI, ONK, SAB and SUC.

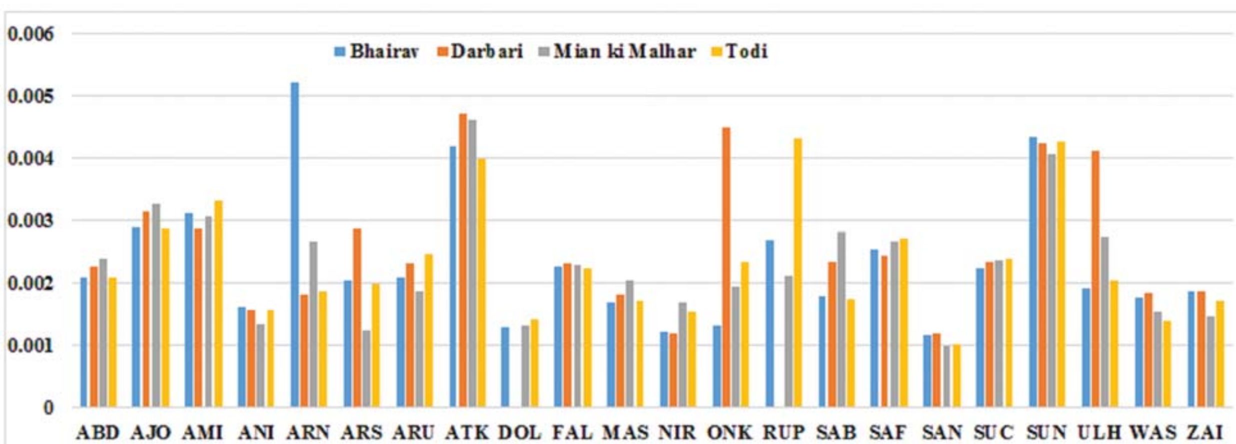


Fig. 2. Frequency of occurrences (along Y axis) for four ragas of twenty-two artists

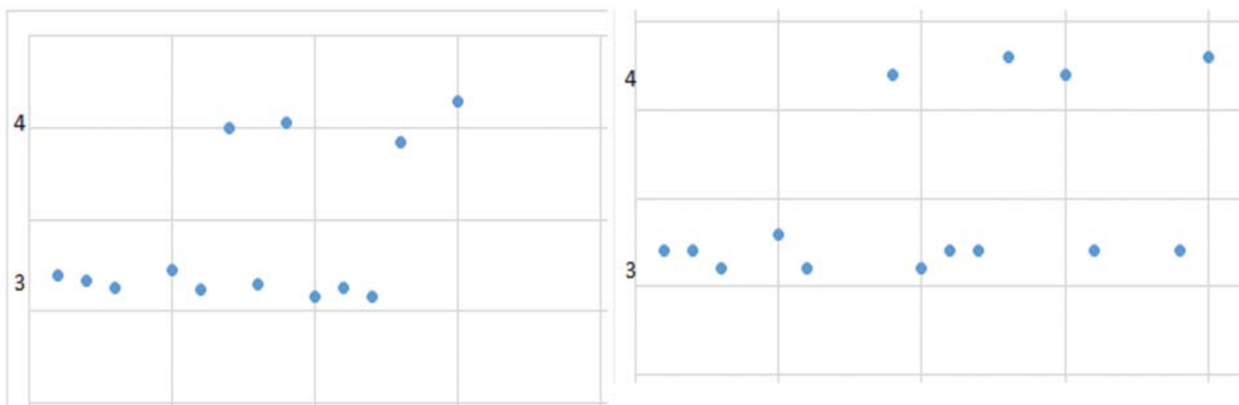
Total number of pause are calculated for each raga of each artist and hence number of such pause per second are measured, called frequency of occurrences as shown in the figure 2. Strikingly uniform frequency of occurrences of silence in between notes is found in all the four signals of the artists SUC, FAL and DOL. Number of such silence per second is an important style stamp of these three artists. Lowest and similar frequency of occurrences of pause are observed in all the raga renderings of the artist SAN. While highest and similar frequency of occurrences of pause are observed in all the raga renderings of the artist SUN. So this can also be considered as a style stamp of the artist SAN and SUN. Very high frequency of pause is found in the renderings of the artist ATK. Frequency of pause is similar in raga darbari kannada and mia ki malhar while raga bhairav and todi shows little difference in frequency of pause for the artist ATK. Similar trend in frequency of occurrences of silence are observed in both the artist ABD, AJO and MAS. Highly similar frequency of occurrences of pause are found in ragas bhairav and todi while very little difference is found in raga darbari kannada and mia ki malhar for the artist ABD, AJO and MAS. Highly similar frequency of occurrences of pause are observed in ragas bhairav and todi while quite high difference is observed in raga darbari kannada and mia ki malhar for the artist ARS, SAB and ULH. Uniformity in frequency of occurrences of silence are observed in the raga bhairav and mia ki malhar while little difference is observed in raga darbari kannada and todi for the artist AMI. Similar frequency of occurrences of silence are observed in the raga bhairav, darbari kannada and todi for the artist ANI. Similar frequency of occurrences of silence are observed in the raga darbari kannada and todi for the artist ARN and ARU but quite high difference is observed in the raga bhairav and mia ki malhar for the artist ARN while little difference is observed for the artist ARU. Uniformity in frequency of occurrences of silence is found in the raga bhairav and darbari kannada but little difference in frequency of occurrences of silence is found in the raga mia ki malhar and todi for the artist NIR, WAS and ZAI. Uniformity in frequency of occurrences of pause is found in the raga bhairav, mia ki malhar and todi for the artist SAF. No trend or uniformity in frequency of occurrences of pause is found in the raga renderings of ONK and RUP.

Table 2. Correlation coefficients between time duration of pause and time duration of phrase

	Correlation coefficient			
	Bhairav	Darbari kannada	Mian ki Malhar	Todi
ABD	0.044	0.312	0.133	0.166
AJO	0.268	0.572	0.125	0.590
AMI	0.066	0.427	0.572	0.195
ANI	0.088	0.546	0.707	0.354
ARN	0.366	0.151	0.105	-0.159
ARS	0.626	0.656	0.563	0.551
ARU	0.212	0.120	0.568	0.214
ATK	0.379	0.389	0.114	0.272
DOL	-0.287		0.162	0.597
FAL	0.504	0.021	0.236	0.540
MAS	0.560	0.344	0.325	0.198
NIR	0.158	0.363	0.211	0.366
ONK	0.576	-0.006	0.388	0.170
RUP	-0.177		0.219	-0.048
SAB	0.235	0.431	0.109	0.286
SAF	0.512	0.220	0.123	0.341
SAN	0.359	0.397	0.165	0.575
SUC	0.281	0.026	0.244	0.306
SUN	0.125	0.078	-0.085	0.208
ULH	0.089	0.083	0.284	0.189
WAS	0.321	0.091	0.154	0.362
ZAI	0.180	0.352	0.432	0.229

No high correlation is found among time duration of pause and time duration of phrase. But moderate correlation is observed in all the four raga renderings of the artist ARS. This is definitely an important finding of the style of this artist. Moderate correlation coefficient is found in two raga signals of the artists AJO, FAL and ANI. Beside these, moderate correlation coefficient is found in one raga rendering only for the artist AMI, ARU, DOL, MAS, ONK SAF and SAN. Rest shows very low correlation among time duration of pause and time duration of phrase.

k nearest neighbour with 4 ragas (k=4) based on the parameter "silence duration" are found among four artists as shown in the figure 3 (A) and k nearest neighbour with 3 ragas (k=3) based on the parameter "silence duration" are found among nine artists as shown in figure 3 (A). k nearest neighbour with 4 ragas (k=4) based on the parameter "silence to note ratio" are found among four artists as shown in the figure 3 (B) and k nearest neighbour with 3 ragas (k=3) based on the parameter "silence to note ratio" are found among ten artists as shown in figure 3 (B). Both the parameters pause durations and silence to note ratio of all the four raga renderings of four artists FAL, ABD, AJO and MAS are closely found in k nearest neighbour. So the silence pattern of these four artists may be considered as their style of singing. Again, silence durations and silence to note ratio of three raga renderings of nine artists AMI, ARN, ARS, ATK, DOL NIR, ZAI, SUN and WAS are closely found in k nearest neighbour.



(A) KNN with pause durations (B) KNN with silence to note ratio

Fig. 3. k nearest neighbour with k=3 and 4

5. CONCLUSION

Similar pattern in frequency of pause is the style identifying parameter for the artist DOL, FAL and SUC. Important findings in the signals of the artist ABD and AJO is that majority of pause is preceded by or followed by meend. Also moderate correlation between time duration of pause and time duration of phrase is a style stamp of the artist ARS. Pause patterns are close in the renderings of the artists FAL, ABD, AJO and MAS.

Hence, analysis of pause can signify certain information about the raga movement. Pause duration, pause to phrase ratio and frequency of occurrences of pause signify the style of presenting a raga by an artist. Difference in pause production among artists is due to their breathing power and creative emotional perspective called improvisation. This study may help the music learners to learn improvisation of a raga.

6. REFERENCES

[1] Brihadaranyaka Upanishad (iv.i.2). translated by Swami Madhavananda with an introduction by Mahamahopadhyaya Prof. S. Kuppuswami Sastri, 1950. published by Swami Yogeshwarananda, Advaita Ashram, Mayavati Almora, Himalayas (3rd edition 1950).

[2] Swami Prajyananda, 1973. Historical Development of Indian Music:, published by - Firma K.L. Mukhopadhyay, Calcutta.

[3] Laxmi Narayan Ghosh, 1975. Geet - Baddyam (part-1):, Pub: Pratap Narayan Ghosh, 1st edition, Calcutta.

[4] Kumar Prasad Mukhopadhaya, 1995. Kudrat Rangi-Birangi:, Ananda Publishers Pvt. Ltd. 1st edition, Calcutta.

[5] Ashok Kumar Datta, Ranjan Sengupta, Nityananda Dey and Dipali Nag, 2006. *Experimental Analysis of Shrutis from Performances in Hindustani Music. Monograph*, Kolkata: ITC Sangeet Research Academy.

[6] Ashok Kumar Datta, Ranjan Sengupta, Nityananda Dey, Dipali Nag and A.A. Mukerjee, 2006. Methodology of note extraction from the song signals, *Proceedings Frontiers of Research on Speech and Music (FRSM-2006),* U P Technical University, Lucknow.

[7] Anirban Patranabis, Kaushik Banerjee, Ranjan Sengupta, Nityananda Dey and Dipak Ghosh, 2012. Search for spectral cues of emotion in hindustani music, *National Symposium on Acoustics (NSA-2012)*, K S R institute for engineering and technology, Erode, Tartist Cl Nadu.

[8] A.K. Datta, S.S. Solanki, R. Sengupta, S. Chakraborty, K. Mahato and A. Patranabis, 2017. *Signal Analysis of Hindustani Classical Music*, Springer Nature, ISBN 978-981-10-3959-1 (eBook).

The Journal of Acoustical Society of India : Vol. 48, No. 1-2, 2021 (pp. 76-81)

Auditory stream segregation in singers and Non-singers : A Comparative Study

Namitha Jain, Zohra Nafees Ghori, Kavitha Janardhan and Prashanth Prabhu

All India Institute of Speech and Hearing, Manasagangothri,
Mysore-570006, Karnataka, India
e-mail: : zohraghori569@gmail.com

[Received: 20-11-2020; Revised: 31-12-2020; Accepted: 31-12-2020]

ABSTRACT

The human brain analyses the complex auditory environment; it perceptually groups the sounds arriving from one source and separates them from competing sources. This phenomenon is called auditory stream segregation. A wide variety of cues help in stream segregation, and Spectral profile analysis is one of them. This study aimed to evaluate and compare auditory stream segregation abilities across singers and non-singers. It was hypothesized that singers would perform better on spectral profiling tasks than non-singers. This study additionally assessed the Spectral profile analysis threshold at different frequencies. The test was administered on 30 singers (with Carnatic music training for at least three years) and 30 non-singers with normal hearing sensitivity. The findings of this study showed that the spectral profiling abilities of singers were not significantly different from the non-singers group. Further, no correlation was found between the profile analysis thresholds and either the duration of Carnatic music training or daily practice. Thus, auditory stream segregation abilities are not significantly enhanced in singers compared to non-singers. However, further research involving electrophysiological tests and a larger group is essential to validate these results.

1. INTRODUCTION

Normal hearing individuals can effortlessly identify the melody of a flute in a musical composition or easily understand someone's speech in a noisy environment. These tasks, though seemingly easy, involves complex processing in the human brain. The brain analyses simultaneous or successive sounds present in the environment and attribute it to a single source or multiple sources. The sounds that arrive from one source in the environment are grouped together perceptually and segregated from competing sounds (such as background noise and conversation). This process is termed as "auditory stream segregation"[1]. There are a variety of cues that help us in segregating or grouping sounds, which are important for auditory streaming. The acoustic cues, such as the temporal proximity of two sounds, the spectral similarity of the successive sounds, are grouped as a single stream. In addition to the acoustic cues (bottom-up cues), there are top-down cues such as attention expectations, and prior exposure plays an important role in auditory stream segregation[2,3]. This study focuses on Spectral profile analysis, which serves as a major cue for stream segregation.

© 2021 Acoustical Society of India

Spectral profile analysis involves recognizing the segregated stream of an auditory scene / a complex auditory environment based on differences in spectral shape. When the components of a complex sound vary in their frequency or intensity together / in harmony, it is perceived as a single stream otherwise perceived as segregated streams. A sound is segregated from the rest of the auditory environment when a component of the complex tone is varied in frequency or intensity, forming an irregular spectral shape. Extensive research in this area has demonstrated that listeners perceive a single perceptual stream or fused image when a group of logarithmically spaced frequencies formed a regular spectral pattern (*e.g.*, a tone composed of 650, 850, 1050, 1250, and 1450 Hz frequencies). When one of these components was altered with respect to frequency or intensity, it induced inharmonicity. This appears as a single bump in the regular spectral pattern and is recognized as arriving from a separate source or simply perceived as separate streams[4]. Literature review shows that mere alteration in intensity does not provide a strong cue for segregation[5,6]. Thus, a listener relies on spectral shape for spectral profile analysis tasks and is independent of the overall level. This ability of our auditory system to detect alterations in spectral shapes can be used to assess spectral profiling abilities[7]. This study utilizes an alteration of the intensity of the middlemost component of complex sound to assess the spectral profiling abilities of participants.

Auditory stream segregation in musicians is of interest because they are extensively trained to perceive fine variations in acoustic stimuli based on amplitude, frequency, and temporal aspects. Musical training provides more auditory experiences. The effects of such increased auditory experiences due to musical training on various psychoacoustic skills have been studied substantially. Musicians are more sensitive to variations in pitch, time, and loudness of the auditory stimuli[8,9], with just noticeable differences being better in musicians than non-musicians. A multitude of evidence suggests that the neuronal coding of fast-changing auditory stimuli is enhanced in musicians[10]. The enhanced neuronal coding due to musical training improves temporal perception abilities[11], speech perception in noise[12] and better fine structure abilities[13]. A positive effect of musical training on stream segregation abilities is reported in the literature [14]. Musicians with greater experience performed better on stream segregation tasks[15]. Spectral profile analysis requires individuals to analyze the spectral composition and identify the stimulus with an inharmonic component. A few studies have attempted to understand auditory streaming abilities in musicians[15], but no studies have analyzed auditory streaming abilities in singers and non-singers. Thus, the current study aimed to understand the effect of Carnatic music training on spectral profile thresholds. A comparison of thresholds across different characteristic frequencies formed the objective. Additionally, years of experience and duration of practice per day were also determined to examine its correlation with spectral profile thresholds.

2. METHODOLOGY

2.1 Participants

The current study adopted a standard group comparison method and recruited 60 individuals within the age range of 15-30 years with no history of ear infection and normal hearing sensitivity. The participants were further divided into two groups; Group 1 composed of 30 singers (mean age= 19.17; S.D= 2.93; 3 males and 27 females) and Group 2 composed of 30 non-singers (mean age= 19.43; S.D= 3.72; 15 males and 15 females). The participants of Group 1 (singers) had at least three years of formal Carnatic music training, and Group 2 (non-singers) received no formal or informal music training. All the participants had pure tone thresholds consistent with normal hearing sensitivity, i.e., within15 dB HL. 'A' type tympanogram and presence of acoustic reflexes (both ipsilateral and contralateral) at all test frequencies indicated normal middle ear status for all participants.

2.2 Procedure

The Audiological test battery was administered on all participants, along with a detailed interview. Participants of Group 1 were asked additional questions, which included the duration of music training and singing practice each day. The audiological test battery included Pure tone audiometry and Immittance evaluation. Pure tone audiometry was conducted on all participants using a calibrated two- channel diagnostic audiometer in sound-treated double room setup. The pure tone average was calculated as an average of four frequencies (500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz). Middle ear status was assessed using Grason Stadler Inc-Tympstar V 2.0 middle ear analyzer. Immittance evaluation included both tympanometry and reflexometry. Tympanograms and acoustic reflexes were obtained at a probe frequency of 226 Hz. The acoustic reflex eliciting stimulus was presented ipsilaterally and contralaterally at various frequencies (500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz) to obtain acoustic reflex thresholds.

Auditory streaming abilities of the participants were assessed using the Profile analysis task in the MATLAB software with the psychoacoustics toolbox. A Profile analysis test was conducted at four frequencies viz. 250 Hz, 500 Hz, 750 Hz, and 1000 Hz. The presentation of different frequencies was randomized for all the participants of the study. In this test, the participant listens to three complex tone stimuli. Among them, two are identical, referred to as standard tones, and the odd one was referred to as variable tone. Each complex tone stimuli have five harmonics. The five harmonics of standard tone have the same amplitude (f0=330-Hz, mi4). The variable tone has a similar harmonic structure with a higher amplitude of the middlemost (third) component of the complex stimuli. This increased amplitude caused irregularity in the spectral shape and was perceived as a different timbre in comparison to the standards. The task of the participant was to identify the odd timbre tone. The subjects were instructed to type the number that produced odd timbre, and the personal computer provided feedback. Every trial, the overall intensity level of both standard and the variable tone was roved randomly within a range of 5 decibels. The onset and offset of stimuli were gated on and off with two 10-ms raised cosine ramps. This experiment was carried out as a 3-alternate forced-choice method. The spectral profile analysis thresholds were calculated as the average of the last four reversals and recorded in decibels. All the stimuli were presented through a personal computer routed through an audiometer, equipped with TDH 39 circum- aural headphones at the most comfortable level of loudness (70dB HL). The complete testing was carried out in a sound-treated double room setup.

3. RESULTS AND DISCUSSION

The comparison of the profile analysis threshold between singers and non-singers showed similar performance across all the frequencies. The mean and standard deviation of the profile analysis threshold across all the frequencies between both groups is shown in figure 1.

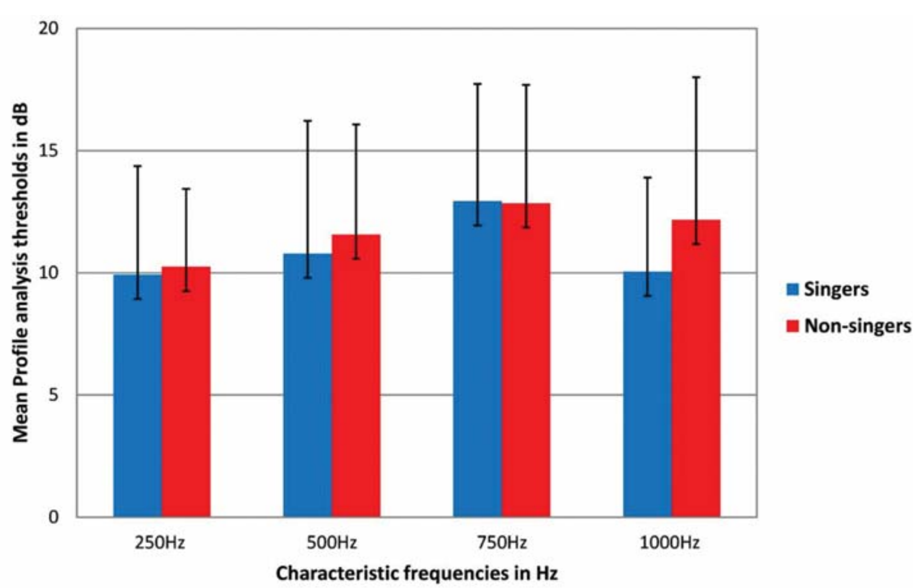


Fig. 1. Shows the mean profile analysis threshold and standard deviation at 250, 500, 750 and 1000Hz frequencies for Non singers and Singers group

The Shapiro Wilk test of normality suggested that the data were normally distributed ($p>0.05$) at all the frequencies, *i.e.*, statistical distribution is normally distributed. Thus, parametric tests like multiple analysis of variance (MANOVA) was done. The profile analysis thresholds at all frequencies (250 Hz, 500 Hz, 750 Hz, and 1000 Hz) were considered as dependent variables between singers and non-singers. A Significant difference (p) was considered at 0.05 for inferential statistical analysis. The results of MANOVA suggested that there is no significant difference [$F (4, 54) = 1.07$, $p>0.05$] between the groups for all the frequencies. In addition, Pearson's product-moment correlation was done to determine if there is any correlation between profile analysis threshold at different frequencies with duration music training and singing practice each day. The results of correlation analysis suggested that there was no significant correlation between profile analysis thresholds and singing practice (years of singing experience and hours of practice per day). The correlation coefficient values with significance level are shown in table 1. The profile analysis thresholds at different frequencies *viz.*, 250 Hz, 500 Hz, 750 Hz, and 1000 Hz against the singing practice in years are shown in figure 2.

Table 1. Correlation co-efficient and significance value for Pearsons correlation analysis

Frequency	Years of singing experience	Hours of practice/day
250 Hz	$r = -0.06$, $p > 0.05$	$r = 0.05$, $p > 0.05$
500 Hz	$r = -0.28$, $p > 0.05$	$r = -0.25$, $p > 0.05$
750 Hz	$r = 0.12$, $p > 0.05$	$r = 0.37$, $p > 0.05$
1000 Hz	$r = -0.44$, $p > 0.05$	$r = 0.05$, $p > 0.05$

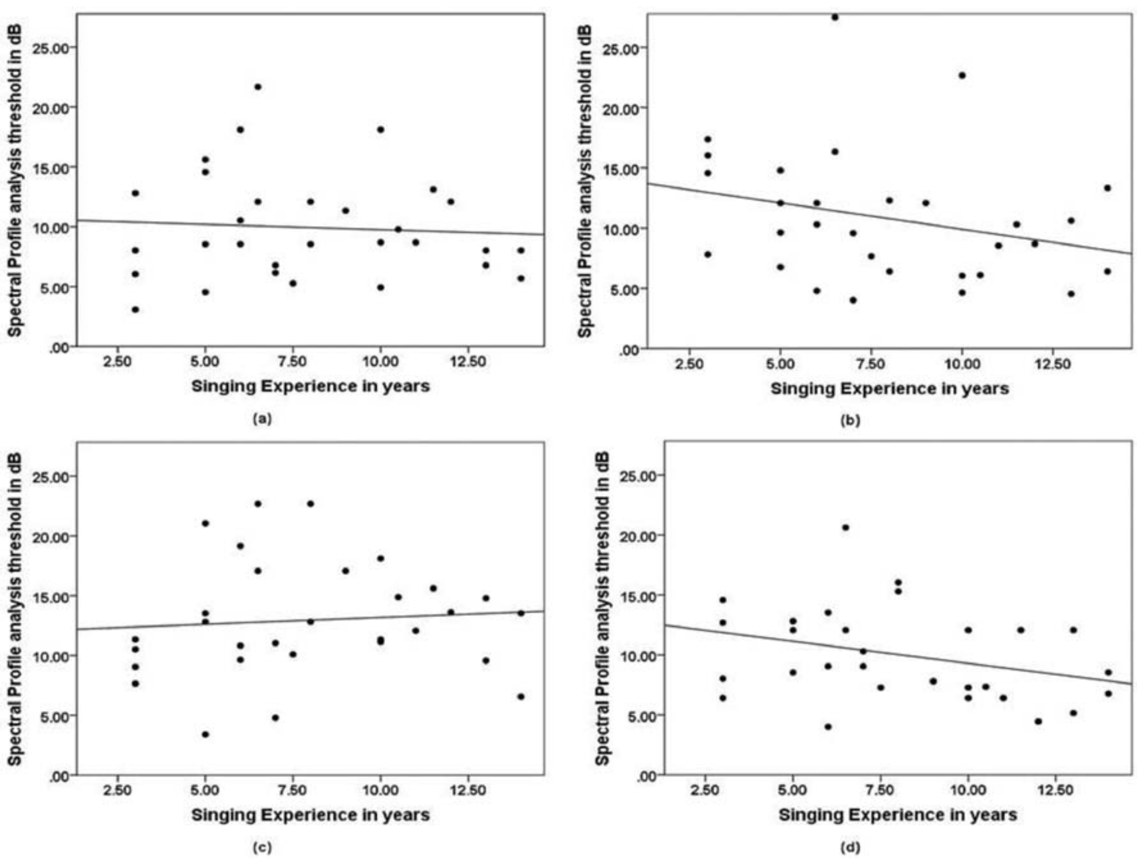


Fig. 2. Spectral profile analysis thresholds as a function of Singing experience in years for participants of group 1 across (a) 250 Hz, (b) 500 Hz, (c) 750 Hz and (d) 1000 Hz characteristic frequencies.

The comparison of spectral profile analysis threshold between singers and non-singers showed no significant difference. The findings of our study also demonstrated that years of singing training and the number of hours of practice per day did not correlate with the profile analysis threshold. This suggests that Carnatic music training does not have an influence on improving auditory stream segregation. The results of the current study are in contrast with previous studies that report that musicians have better auditory stream segregation skills[9, 15]. However, the previous studies were on instrumental musicians and not on singers. The studies also report that the auditory stream segregation abilities vary depending upon the type of instrument played by the musician[16]. In addition, it is also reported that top-down influences, such as attention, can also play a major role in stream segregation. The spectral profile analysis task is behavioral in nature, and it is significantly influenced by attention. We found in our study that the participant's attention altered the profile analysis threshold significantly. The performance also improved with practice. Thus, the result of a behavioral test is difficult to interpret because of individual differences in the amount of attention during the task.

The results of the study highlight that there is no significant enhancement of auditory stream segregation abilities among singers compared to non-singers. The study was carried out in singers who were adolescents and adults. Thus, spectral profile analysis abilities might have improved with exposure and reached an asymptote. Hence, if a similar study is carried out on children, it would shed light on the development of auditory stream segregation with age. The result was obtained only using the spectral profile analysis task, and further studies on other measures of auditory stream segregation are essential. It is also important to analyze the effect of singers with and without instrumental music training to know the effect of instrumental music on the profile analysis threshold. In addition, further studies are also essential comparing singers with Carnatic, Hindustani, and Western Music training to know the differences. The present study is just a preliminary report on auditory stream segregation in singers. Further studies using electrophysiological tests, a larger sample size with more variables, and the effect of different styles of musical training should be considered to understand any differences in auditory stream segregation abilities with or without music training.

4. CONCLUSION

The present study aimed to evaluate behavioral differences between singers and non-singers in auditory stream segregation. Spectral profiling, a test of auditory stream segregation, was assessed at various characteristic frequencies, and thresholds were determined. The findings of the present study show that spectral profiling abilities across singers and non-singers had no significant difference. Hence, it can be inferred that there is no significant effect of Carnatic music training on stream segregation abilities. In addition, it was also found that the years of Carnatic music training and duration of daily practice did not correlate with profile analysis thresholds. In conclusion, auditory stream segregation is not enhanced in singers compared to non-singers. However, further studies with electrophysiological tests and a controlled environment are essential for validation of the results.

5. REFERENCES

[1] A.S. Bregman, 1994. "Auditory scene analysis: The perceptual organization of sound".

[2] N. Marrone, C.R. Mason and G. Kidd, 2008. "Evaluating the benefit of hearing aids in solving the cocktail party problem.," *Trends Amplif.*, **12**(4), 300-15, doi: 10.1177/1084713808325880.

[3] B.G. Shinn-Cunningham and V. Best, 2008. "Selective Attention in Normal and Impaired Hearing," *Trends Amplif.*, **12**(4), 283-299, doi: 10.1177/1084713808325306.

[4] J.M. Brunstrom and B. Roberts, 1998. "Profiling the perceptual suppression of partials in periodic complex tones: Further evidence for a harmonic template," *J. Acoust. Soc. Am.*, **104**(6), 3511-3519.

[5] N.J. Versfeld and A.J. Houtsma, 1991. "Perception of spectral changes in multi-tone complexes". *The Quarterly Journal of Experimental Psychology*, **43**(3), 459-479.

[6] J.J. Lentz, 2005. "Profile analysis: The effects of rove on sparse spectra". *The Journal of the Acoustical Society of America*, **118**(5), 2794-2797.

[7] D.M. Green, 1983. Profile Analysis: Auditory Intensity Discrimination. Oxford: *Oxford University Press.*

[8] J. Marozeau, H. Innes-Brown and P.J. Blamey, 2013. "The effect of timbre and loudness on melody segregation," *Music Percept. An Interdiscip. J.*, **30**(3), 259-274.

[9] J. Marozeau, H. Innes-Brown, D.B. Grayden, A.N. Burkitt and P.J. Blamey, 2010. "The effect of visual cues on auditory stream segregation in musicians and non-musicians," *PLoS One*, **5**(6), e11297.

[10] D.L. Strait, S. O'connell, A. Parbery-Clark and N. Kraus, 2013. "Musicians' enhanced neural differentiation of speech sounds arises early in life: developmental evidence from ages 3 to 30," *Cereb. Cortex*, **24**(9), 2512-2521.

[11] T. Rammsayer and E. Altenmüller, 2006. "Temporal information processing in musicians and nonmusicians," *Music Percept. An Interdiscip. J.*, **24**(1), 37-48.

[12] C. Jain, H. Mohamed and A.U. Kumar, 2015. "The effect of short-term musical training on speech perception in noise," *Audiol. Res.*, **5**(1), 111.

[13] S.K. Mishra, M.R. Panda and S. Raj, 2015. "Influence of musical training on sensitivity to temporal fine structure," *Int. J. Audiol.*, **54**(4), 220-226.

[14] B.R. Zendel and C. Alain, 2009. "Concurrent sound segregation is enhanced in musicians," *J. Cogn. Neurosci.*, **21**(8), 1488-1498.

[15] N. Johnson, A.M. Shiju, A. Parmar and P. Prabhu, 2020. "Evaluation of Auditory Stream Segregation in Musicians and Nonmusicians," *Int. Arch. Otorhinolaryngol.*

[16] S. Sauve, L. Stewart and M. Pearce, 2014. "The effect of musical training on auditory grouping".

The Journal of Acoustical Society of India : Vol. 48, No. 1-2, 2021 (pp. 82-90)

Perception of ornamentation in Hindustani classical music

Junmoni Borgohain[1*], Raju Mullick[2], Gouri Karambelkar[2], Priyadarshi Patnaik[1,3], Joy Sen[4] and Damodar Suar[1]

[1]*Department of Humanities and Social Sciences,*
Indian Institute of Technology Kharagpur, Kharagpur-721302
[2]*Advanced Technology Development Centre,*
Indian Institute of Technology Kharagpur, Kharagpur-721302
[3]*Rekhi Centre of Excellence for the Science of Happiness,*
Indian Institute of Technology Kharagpur, Kharagpur-721302
[4]*School of Architecture and Regional Planning,*
Indian Institute of Technology Kharagpur, Kharagpur-721302
e-mail: : jun.borgohain12@gmail.com

[Received: 30-11-2020; Revised: 28-12-2020; Accepted: 28-12-2020]

ABSTRACT

In Indian aesthetic tradition - especially poetry and music - *alamkara* (lit. ornamentation) or embellishment is considered a key aesthetic component that 'embellishes' or beautifies a piece. However, not much research has been dedicated to examine the role *alamkaras* play in Hindustani classical music. In this paper we ask the question, do *alamkaras* merely beautify a piece or they also perform specific functions to influence how music is perceived in terms of emotions and other expressive features? One hundred and five participants listened to and judged four raga clips, two happy and two sad, rendered using three *alamkaras* - *meand, murki,* and *khatka*- in their nonarticulated forms. They rated the clips on emotions and perceptual attributes conveyed by them. Results suggest that ornamentations play a key role not only in the intensification of primary emotions but also in variations in secondary associated emotions, which change the entire color of the *raga* rendition. They also suggest that expressive features such as harshness-softness, aggressiveness-mildness, jerkiness-smoothness, strange-familiar and Masculine-feminine get significantly modified by them. Findings are relevant in the context of Indian musicology since they throw light of the specific roles key *alamkaras* play in coloring emotions and associated expressive features.

1. INTRODUCTION

The underlying aim of Hindustani Classical music (HCM) is to evoke aesthetic emotions *(rasas)* and induce an aesthetically pleasing experience[1]. *Ragas* are melodic configurations of a group of notes rendered based on the traditionally accepted rules; improvised, embellished, and expanded. Traditional

© 2021 Acoustical Society of India

knowledge suggests that the traditionally attributed emotions for *ragas* are communicated through features woven within a *raga*. These include, for example, scales used *(thaat)* sequence of notes *(chalan)*, repetition of specific clusters of notes *(pakad or mukhyang)*, emphasis on dominant notes *(Vadi-samvadi)* and specific touches, glissando *(meand),* or the way that notes are touched continuously between three *(khatka),* or four notes *(murki)* during the rendition of a *raga*. The last three components, namely, *meand, khatka,* and *murki*, interestingly, also come under the category of *alamkaras*.

The swaras or musical notes are rendered with *alamkaras* to bring out the aesthetic essence of a *raga*[2]. In actual practice, ornamentations cited above are used organically in the exploration of the relationship of certain notes or motifs specific to each *raga* to bring out its essence. Traditionally, *alamkaras* are used specifically to enhance the pleasantness of a musical piece. They are believed to perform decorative or beautifying functions without affecting the structure, emotion, or meaning of a *raga* in any fundamental way. However, some recent practitioners of music suggest that *alamkaras* perform more fundamental functions such as defining the very identifying features of a *raga* and establishing its mood. Evidence from western music has investigated the role of ornamentation and found that it performs manifold functions ranging from augmentation of expression and meaning to the melody, and expression of distinct emotions among others[3,4]. Apart from that, recent findings report that ornamentation assists the performer in conveying the expression intended in the piece and the interpretation of the underlying musical structure[5].

Alamkaras are an integral part of the melodic structure of a *raga*. A musical note qualifies to be termed as 'musical' if it is rendered with the appropriate *alamkara*. However, less attention has been devoted to the communication of emotions and their related qualities by *alamkaras*. This study focuses on understanding the emotional connotations and perception of attributes in *raga* clips concerning three specific *alamkaras* - *meand, murki,* and *khatka*- when they are used in their non-essential contexts, *i.e.*, where they are not required to establish the identity of a *raga*. The HCM artist has freedom to select an *alamkara* to be applied or change the *alamkara* applied according to the requirement of the *raga-bhava* or her / his own perception / interpretation of that *raga*.

2. HINDUSTANI CLASSICAL MUSIC AND AESTHETIC EMOTION *(rasa)*

The word *raga* is defined as *'ranjayateitiraagah'* means *'raga* is which pleases the mind'. Further, its general lexical meaning is also emotion, color, and so on. It may be defined broadly as a melodic scheme, characterized by a definite scale or notes (alphabet), order of sequence of these notes (spelling and syntax), melodic phrases (words and sentences), pauses and states (punctuation), and tonal graces (accent)[6]. According to both ancient text as well as the living tradition, the essence of a *raga* in HCM is the exploration, expansion, and intensification of a cluster of compatible emotions that finally lead to the experience of one dominant aesthetic emotion in the viewer - this is known as *rasa*[1]. *Rasa* theory, which originated with *Natya Shastra*, suggests that any work of art, when experienced to its fullest, leads to an aesthetic state of pleasure which is linked to certain emotions, but at a higher level. It also states that the ultimate aim of a *raga* is to bring about the blossoming of emotions in the competent listener. This is brought about because of the various attributes / features of a *raga* (it's *vibhavas*) that lead to the generation of emotive responses (*anubhavas* and *sancharibhavas*) which finally lead to a distinctive kind of emotive response-*rasa*[7,8]. Among the various devices *vibhav*as that lead to the generation of emotions, *alamkaras* are identified as those which perform the role of beautifying a *raga* or making it graceful.

2.1 *Alamkars* in HCM *ragasangeet*

Alamkaras : In Indian classical music, staccato notes or isolated notes are not frequently used. While rendering a *raga*, each note is linked to its preceding or succeeding using one of the *alamkaras*. *Alamkara* in the literal sense means jewellery and in the *Natyashastra* it is said that 'a melody without ornament is like a night without moon, a river without water, a vine without flowers, or a woman without jewels' (Nat. Sh. 29,75). *Alamkaras* or 'tonal graces' can either focus on the interplay between notes or the treatment of notes without affecting their order. Hence, they cannot be considered extrinsic or accidental to a *raga*[2]. Each *raga* acquires its uniqueness because of not only its musical notes but also since the latter is aesthetically rendered, accompanying graces. *Alamkaras* also perform other functions such as differentiating one *raga* from the other. For instance, *raga Des* and *Tilak Kamod* have the same notes but the *meand* is applied to different notes in the *raga,* with help of which one can differentiate between the two *ragas* in a single phrase. R P or G N S (with *meand*) in *raga Des* and P S or S P in *raga Tilak Kamod,* use the *meand* but the phrases and thus the *raga* chalan is different.

They are of two basic categories : *varnalamkar* and *sabdalamkar*. *Varnalamkaras* explore and use sequences of notes but later it changed to the quality of intonation[1] whereas *sabdaalamkaras* focus on the sonic production rendered by the voice or any instrument. The former focuses on the structural aspect and the latter is related to the aesthetic aspect of a *raga*. *Sabdalamkaras* comprise the tonal ornamentations (*khatka, murki, Meand, kan, etc.*) known as *alamkaras* or ornamentations. The other group of *shabdalamkaras* is termed as embellishments such as *alap, tana, behelava, etc*. The three alamkaras used in this study are, *meand, khatka,* and *murki*.

2.1.1. Meand : *Meand* has its root in the word *meadam* which means 'in a low tone' or 'softly'. It is a soft glide between two notes, or between two notes in different octaves. *Meand* is one of the toughest ornaments in Hindustani music because its behavior between two notes depends completely on the rules of the *Raga*.

2.1.2. Khatka : *Khatka* is derived from the verb *'khatakna'* which means to create a sharp clashing sound. In music, it is a melodic embellishment wherein a group of notes is quickly and forcefully rendered focusing on the important note of that cluster[1]. The main note is embellished with a touch of the next higher note and then with that of the next lower note, or vice versa.

2.1.3. Murki : *Murki* is derived from the Hindi verb *'murakana'* which means a turn or a quick twist. *Murki* refers to a quick circulatory movement of notes. It is rendered in a cluster around the main note, but instead of involving four notes, it does so with only three notes.

3. PERCEPTUAL STUDIES ON ORNAMENTATION

In the last four decades, researchers have attempted to understand how music is perceived and experienced by listeners. Researchers have illustrated models to show the relationship between the performer-adapted strategies to convey emotions and the affective experience of listeners. One such model is the Lens model that is an extensively used framework to investigate the affect judgment process. The performer (encoder) uses certain cues to express affect in a musical piece that listeners (decoder) utilize to attribute emotions[9]. There are also assertions that emotion in music is conveyed through both culture-specific (for example- mode) and perceptual cues (in the form of psychophysical dimensions of music). So, performers consciously or intuitively might draw upon them to express emotion in music. Listeners, in turn, might attend to either of these sources for emotional meaning[10,11]. Some of the acoustic cues suggestive of evoking emotions in listeners are tempo, mode, harmony, tonality pitch, interval, rhythm, sound level, timbre, timing, articulation, accents on specific notes, tone attacks and decay, and vibrato[5,12]. In the context of HCM, explorations were made to study the role of tonic intervals, rhythmic regularity, and tempo in the experience of emotions[8,13,14].

The use of ornamentation is also regarded as one of the components linked with the emotional expressiveness of music. Traditional knowledge incidentally also indicates that unless *alamkaras* are used, the *ragas* will not possess the intended color or flavour. However, the literature on investigating the influence of the use of distinctive *alamkaras* and their affective content is negligible[15]. In the western context, ornamentations were used in Baroque music and their use flourished in that era. A study was conducted to examine the influence of various performer-adapted ornamentations on the perception of emotions[4]. Two professional musicians performed three melodies to express happiness, love, sadness, and anger in two forms - mild and intense. Results showed that listeners were successfully able to decode the emotions in the performance. Ornamentation, therefore, can be assumed to be musical cues that assist in decoding the expressiveness of a music clip.

In the present study, we ask the question of whether *alamkaras* are mere embellishments or play a significant role in the elucidation of a *raga* in terms of emotions conveyed and expressive features evoked. We aimed to investigate if the use of *alamkaras* can communicate emotions and other perceptual attributes. We expected that ornamentation as a cue and an organic feature of a *raga* can reinforce and manipulate the emotive content of the *raga.* We contended that there will be a change in perceptual attributes mostly accountable to feature changes in the clips.

4. METHOD

4.1 *Participants*

One hundred five participants who were pursuing undergraduate studies at the Indian Institute of Technology, Kharagpur participated in the study 38% were females and 62% were males with an age range from17-21 yrs. (M=18.3, SD=0.7). Amongst the participants, four of them reported musical training in Hindustani and western classical music that ranged from 0-14 years.

4.2 *Stimuli used*

The four *ragas* selected for the study were traditionally attributed as happy and sad *ragas* rendered in *shuddha* and *komal swaras*. The happy and sad *ragas* were *Shree, Hansadhwani,* and *Des, Marwa* respectively. A vocalist trained in the Hindustani Classical music tradition rendered them in slow tempo *(alaap)* for 15-20 seconds. Three *alamkaras- meand, murki* and *khatka* were added to each of the *raga* clips. A set of simple happy and sad *raga* clips without using any of the *alamkaras,* using staccato notes) were also rendered that were used for comparison with the ornamented clips.

4.3 *Procedure*

The study was conducted using a web-interface. Participants were notified earlier and instructed to have headphones available on their systems. Web-links were sent to them through email and they were given a window of 48 hours to complete the task as a part of their class assignment. Before starting the task, complete instruction for the survey was given on the first page. After they were ready, the start survey button redirected them to the next page of the interface where the primary task was outlined. Respondents were asked to listen to both the clips and compare the second clip (ornamented added) with the first (non-ornamented). After listening to the clips, the participants had to rate the clip on a semantic-differential scale (SDS). The bipolar adjectives in the SDS ranged from a scale of -1 to -3 and +1 to +3. Participants also rated the emotion perceived in the clip and the intensity ranging from a scale of 1-5 whether they find it diminished or increased in the second clip.

4.4 *Data Analysis*

The Pearson Chi-square test was used to analyze the statistical difference in the perception of emotions with three *alamkaras*. The One-way ANOVA was utilized to examine if statistical difference existed in the rating of perceptual attributes when *alamkaras* were added.

5. RESULTS AND DISCUSSION

5.1 *Emotions perceived in the ragas*

Chi square was calculated to determine the significant differences in the perception of emotions in the *ragas* when *alamkaras* were added to them. For *raga Des*, when rendered using *meand*, 'happy' (32.1%) was the dominant emotion in the clip followed by 'calm' (31.1%). The same *raga* rendered in *murki* and *khatka* were also perceived as happy. However, ratings of happiness were more in *murki* (45.6%) than in *khatka* (39.6%). The secondary emotion perceived in both the ornamentations was 'exciting'. The chi square value also yielded significant results showing that the emotions ratings in three clips in different ornaments were significantly different (x^2= 67.85, p <.001).

Raga Hansadhwani was perceived as happy dominantly when *meand, murki* and *khatka* were utilized. Results showed that 'calm' (28.3%) was perceived higher in *meand* next to happy (30.4%) as compared to *murki* and *khatka*. 'Exciting' was perceived significantly in *murki* (20.7%) and *khatka* (25.8%) as a secondary emotion. The chi square statistics also showed a significant difference in the perceived emotions in *raga Hansadhwani* (x^2=54.57, p <.001)

Table 1. Emotions perceived in happy *ragas*

Raga	***Alamkara***	**Anger**	**Calm**	**Exciting**	**Fear**	**Happy**	**Romance**	**Sad**	**Wonder**	**Chi-Square**
		%								
Des	*Meand*	8.5	31.1	0.9	0.0	32.1	1.9	19.8	5.7	67.85***
Des	*Murki*	2.9	8.7	20.4	2.9	45.6	1.9	14.6	2.9	
Des	*Khatka*	13.9	5.0	18.8	5.0	39.6	5.0	10.9	2.0	
Hansadhwani	*Meand*	9.8	28.3	0.0	1.1	30.4	2.2	25.0	3.3	54.57***
Hansadhwani	*Murki*	3.3	12.0	20.7	4.3	41.3	2.2	13.0	3.3	
Hansadhwani	*Khatka*	0.0	18.3	25.8	0.0	41.9	0.0	11.8	2.2	

Note. ***p<.001

Descriptive statistics revealed that the dominant emotion in *Marwa* and the three *variants, meand, murki* and *khatka* was sadness. Calmness (23.9%) was perceived significantly higher in *meand* after sadness than *murki* and *khatka*. In the other two *alamkaras* of *Marwa*, anger received second highest ratings after sadness. Similarly, in Shree, the dominant emotion in the three stylistic variations was sadness. But sadness was perceived more in *murki* (58.1%) than *khatka* (45%) and *meand* (50.5%). The secondary emotions for *meand* and *murki* were anger. However, fear was the most rated emotion after sadness in *khatka*.

Table 2. Emotions perceived in sad *ragas*

Raga	***Alamkara***	**Anger**	**Calm**	**Exciting**	**Fear**	**Happy**	**Romance**	**Sad**	**Wonder**	**Chi Square**
		%								
Marwa	*Meand*	7.60	23.9	2.2	1.1	12.0	4.3	45.7	3.3	50.77***
Marwa	*Murki*	20.4	0.0	9.2	7.1	11.2	0.0	48.0	4.1	
Marwa	*Khatka*	17.2	10.8	2.2	12.9	10.8	2.2	41.9	2.2	
Shree	*Meand*	20.8	7.9	1.0	5.9	9.9	1.0	50.5	3.0	13.03***
Shree	*Murki*	18.1	0.0	2.9	7.6	10.5	1.0	58.1	1.9	
Shree	*Khatka*	19.0	7.0	2.0	22.0	5.0	0.0	45.0	0.0	

Note. ***p<.001

The results of the study are very insightful since it tells us that *alamkaras* do play an important role in modifying the perception of emotions. The happy *ragas* were perceived dominantly as happy in the three ornamentations and similar was in the case of sad *ragas*. However, the underlying secondary emotions in these *ragas* varied. For example, when *meand* was applied in both *Des* and *Hansadhwani* calmness was perceived. Similar was the case with *Marwa* which is traditionally attributed as a sad *raga*, calmness was perceived as a subdominant emotion in *meand*. However, in *Shree-meand* anger was the subdominant emotion. When *murki* and *khatka* were applied to the happy *ragas*, excitement was perceived as the secondary emotion, whereas anger was perceived secondary to sadness in *murki* and *khatka* in *Marwa*. However, the difference was observed in the choice of subdominant emotions in *Shree, Shree* in *murki* was rated as angry and *khatka* as fearful. The emotions perceived show a significant change when *alamkaras* were applied suggesting that *alamkaras* have the potential to influence the affective content of a *raga*.

The difference in emotion perception can be attributed to the fluid nature of music indicating that multiple emotions can be expressed within a single musical passage. The literature is evident with instances suggesting the presence of mixed emotions or a blend of emotions in music[16]. Most of them term them as secondary emotions that are based on primary emotions[17]. The same author proposed a model where emotions were arranged with the intense emotions at the top of the cone and the lesser intense ones on the vertical plane of the cone. For example, emotion pride involved a mixture of happiness and anger. The dimensional model has instances of co-occurrences of mixed emotions and often they are of opposite valences (For example, happiness and anger). Similarly, the discrete model also encourages tapping of the emotional valences of different emotional states[18].

A study by Hunter, Schellenberg, & Schimmack[19] found that mixed cues evoked both happiness and sadness in a music clip. About the choice of emotions in our study, the usage of *alamkaras* mediated the perception of the emotions in the clips. In the Classical tradition, the dominant emotions also known as *Sthyayi Bhava* remain intact whereas the *sanchari bhavas* are the fleeting emotions. They possess the ability to moderate the affect, assuming that they can communicate *anubhavas*. These are small acts within the music which together with *bibhavas* lead to *bhavas* and give an integral experience of the *raga*. In the context of our study, the dominant emotions of the traditionally attributed *ragas* had a high resemblance. However, the subdominant emotions as found in our study showed the prevalence of *sanchari bhavas*. Their presence instigated cognitive appraisal of the clip which led to choosing multiple basic emotions or mixed emotions in the clips.

5.2 *Attributes perceived in the ragas*

The perception of the attributes of the *ragas* in the ornamentations was assessed using one-way ANOVA. A higher mean score on the attributes indicated the presence of positive features whereas a lower mean score indicated the presence of the negative features in the *raga*. In *raga Des* significant differences were observed in jerky-smooth (F(2,307) =7.09, p< 0.001) in the three artistic variations. The Tukey post hoc test revealed that perception of jerkiness was more in *murki* (M= 2.38, SD= .68, p <.001) than *khatka* (M= 3.29, SD=1.7, p<.001) and *meand* (M= 4.1, SD= 1.4, p<.001).

In *raga Hansadhwani*, aggressive-mild (F(2,274)=3.97 p=<0.05) and harsh-soft (F (2,274) =4.79, p=<0.05) were perceived significantly different in the three ornamentations. The mean score on the attribute jerky-smooth showed that the *raga* was significantly perceived jerkier in *murki* (M=3.30, SD=1.5, p< .001) and *khatka* (M= 3.52, SD=1.75, p< .001)as compared to *meand* (M= 4.21, SD=1.26). In the harsh-soft feature, *meand* (M=4.82, SD=1.4) was perceived softer than *khatka* (M=4.23, SD=1.3 p< .001) and *murki* (M= 4.70, SD=1.3, p< .001) as shown in Fig. 1.

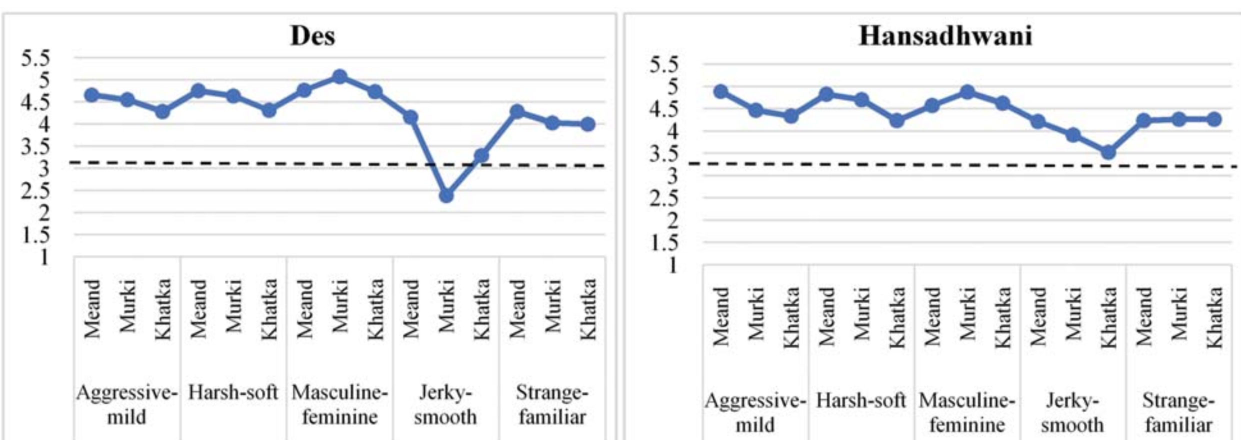


Fig. 1. Ratings of attributes in *Des* and *Hansadhwani*

Significant differences were observed in the all the attributes in *Marwa* except in case of strange-familiar (F(2,280) =1.55, p=<0.05). Post hoc tests revealed that *Marwa* in *meand* was perceived significantly as mild (M=4.78, SD=1.3) than *murki* (M=3.64, SD=1.5) and *khatka* (M=3.61, SD=1.3). *Meand* was also judged softer (M=4.56, SD=1.3) than *murki* (M=3.96, SD=1.3) and *khatka* (M=3.76, SD=1.2). The feature of smoothness was also high in *meand* (M=4.12, SD=1.3) than *murki* (M=3.31, SD=1.4) and *khatka* (M=3.39, SD=1.4). *Meand* was again perceived as more feminine (M=4.66, SD-=1.3) than *khatka* (M=4.15, SD=1.1).

In *Shree* two attributes showed significant differences aggressive-mild (F(2,303) = 2.94, p=<0.05) and harsh-soft (F(2,303) =3.48, p=<0.03). *Shree* was also perceived feminine in all the forms of *alamkaras*. The ratings in all the attributes were below the average rating of 4 showing a dominance of aggression, harshness, jerkiness and strange in the three ornamentations as shown in Fig. 2.

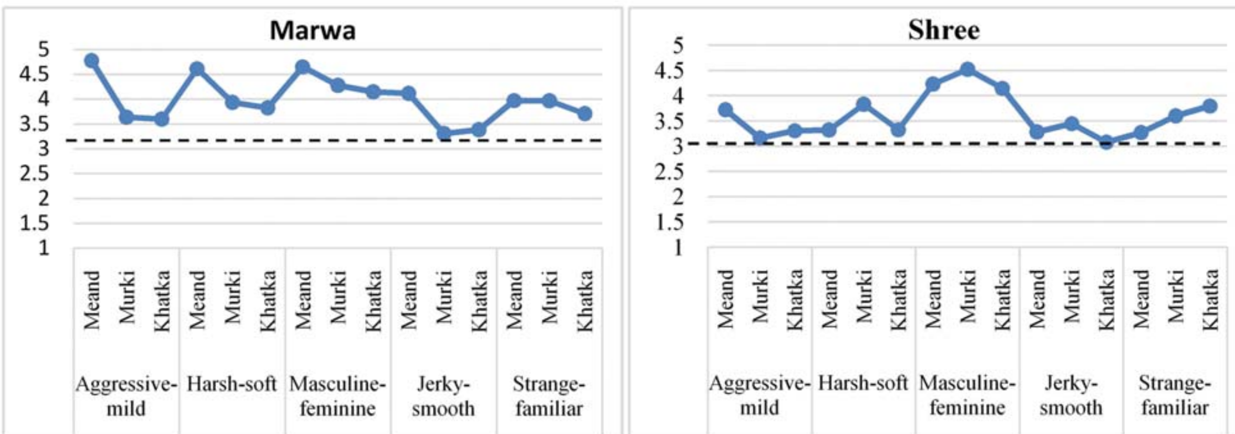


Fig. 2. Ratings of attributes in *Marwa* and *Shree*

The attribution of adjectives to the *raga* clips also changed with designated *alamkaras*. The results suggest that stylistic patterns do influence the way the *ragas* are perceived. Some of the attributes in *Des raga* were perceived above the average value of 4 indicating a positive inclination towards the attributes. They were judged as mild, soft, feminine, and familiar. Exceptions were observed in *khatka* being jerkier than *murki* and *meand*. Similarly, in *Hansadhwani*, all the attributes had a positive orientation except thesame *raga* in *murki* leaning towards the jerkier characteristic.

Distinctive trends were also observed in the sad *ragas*. *Marwa* rendered with *murki* and *khatka* was rated less than the threshold level of 4 suggesting the perception of attributes such as aggressiveness, jerkiness, strangeness, and harshness in the clip. The exception was only with *meand* which was rated low in the above-mentioned attributes. *Raga Shree* however was an exception as it was perceived high in aggressiveness, jerkiness, strangeness, and harshness in all the three *alamkaras*. Nevertheless, both the sad *ragas* were judged as feminine. Generally, it was observed that in *meand*, barring *Shree*, in all the other cases, the *raga* was softened, made mild, feminine and sounded more familiar (see Fig. 1 and 2). In case of *murki* and *Khatka*, it enhanced the elements of aggressiveness, harshness, and jerkiness for all the *ragas*.

The variability of changes observed in both the happy and sad *ragas* can be attributed to the emotional attributes of the *ragas*, the first two being positive (happy and calm) *ragas* and the other two being negative (sad and angry) *ragas*. In that case, it is also clear that *alamkaras* can perform discriminatory functions and enhances the *"raganess"* of the *raga*. *Alamkars* not only provide the beauty & exoticness to HCM but also characterizes and differentiates the *ragas*[20] and their traits in different genres like *Khayal*, *Thumri* or *Dhrupad*.

The variation in the choice of features for the two sets of happy and sad clips confirms the assertion that music is a denotative language[21]. According to Wieczorkowska *et al.*[8] while responding to affective content expressed by Indian *ragas*, a listener describes real-world events. While the feminine aspect of the *ragas* was consistent through the various clips, masculinity was rarely rated highly. It shows the dependence of the interpretation of meaning on musical context which is gendered as suggested by many authors[3,5,22,23]. In fact, the gestures and structures employed within a composition possess direct relationships to traits, qualities, or behaviours perceived to be characteristic of either the masculine or feminine directions of gender[24]. Nevertheless, all the *ragas* in our study were traditionally attributed as male *ragas*, however, we do not see any correspondence between the clips and feature selected.

6. CONCLUSION

In this study, we examined the impact of three *alamkaras* on the perception of emotions and attributes when used with happy and sad *ragas*. We found in our study that the usage of *alamkaras* affected the perception of emotions ranging from positive-negative valence along with changes in secondary emotions evoked. The ratings of expressive features also varied with the use of *alamkaras* suggesting the listeners were sensitive to cues that were embedded in the music clips. The study has certain limitations. Future studies could concentrate on the mapping of emotions for various clips, and more detailed off-line studies, studies with different categories of respondents (musicians, music students as opposed to lay listeners) need to be conducted to throw more light on specific roles *alamkaras* play across *ragas* and different HCM genres.

7. REFERENCES

[1] A. Ranade, 2006. *Music contexts: A concise dictionary of Hindustani music*. New Delhi and Chicago: Promila and Co. Publishers in association with Bibliophile South Asia, p. 254

[2] S.K. Pudaruth, 2016. A reflection on the aesthetics of indian music, with special reference to hindustani raga-sangita. *Sage Open*, **6**(4), 2158244016674512.

[3] P.N. Juslin and R. Timmers, 2010. "Expression and communication of emotion in music performance," in *A Handbook of Music and Emotion*, eds P.N. Juslin and J.A. Sloboda (Oxford, UK: Oxford University Press), pp. 493-492.

[4] R. Timmers and R. Ashley, 2007. Emotional ornamentation in performances of a Handel sonata. *Music Perception*, **25**(2), 117-134.

[5] P.N. Juslin and P. Laukka, 2003. Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* **129**, 770-814. doi: 10.1037/0033-2909.129.5.770

[6] B.C. Deva (Ed.), 1992. Introduction to Indian Music. Publications Division Ministry of Information & Broadcasting.

[7] B.C. Deva, 1973. Classification of Indian Musical Instruments. *Journal of the Indian Musicological Society*, **4**(1), 33.

[8] A.A. Wieczorkowska, A.K. Datta, R. Sengupta, N. Dey and B. Mukherjee, 2010. On search for emotion in Hindusthani vocal music. *In Advances in music information retrieval*. Springer Berlin Heidelberg. pp. 285-304.

[9] E. Brunswik, 1956. *Perception and the representative design of psychological experiments*. Univ. of California Press.

[10] E. Bigand, S. Filipic and P. Lalitte, 2005a. The time course of emotional responses to music. *Annals of the New York Academy of Sciences*, **1060**(1), 429-437.

[11] L.L. Balkwill and W.F. Thompson, 1999. A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. Music Perception: *An Interdisciplinary Journal*, **171**, 43-64.

[12] P.N. Juslin and P. Laukka, 2004. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research,* **333**, 217-238.

[13] S. Hegde, J. Aucouturier, B. Ramanujam and E. Bigand, 2012. *Variations in emotional experience during phases of elaboration of North Indian Raga performance. Proceedings of the 12th International Conference on Music Perception and Cognition ICMPC and 8th Triennial Conference of the European Society for the Cognitive Sciences of Musi*c ESCOM, **1**, 412-413.

[14] A. Mathur, S.H. Vijayakumar, B. Chakrabarti and N.C. Singh, 2015. Emotional responses to Hindustani raga music: The role of musical structure. *Frontiers in Psychology*, 6.

[15] A. Mahajan, 2010. *Ragas* in Hindustani music: Conceptual aspects. New Delhi: Gyan Publishing House.

[16] I. Cross, 2008. Musicality and the human capacity for culture. *Musicae Scientiae*, **12(1_suppl)**, 147-167.

[17] R. Plutchik, 1994. The psychology and biology of emotion. New York: Harper-Collins College Publishers.

[18] K. Vansteelandt, I. Van Mechelen and J.B. Nezlek, 2005. The co-occurrence of emotions in daily life: A multilevel approach. *Journal of Research in Personality*, **39**(3), 325-335.

[19] P.G. Hunter, E.G. Schellenberg and U. Schimmack, 2010. Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity and the Arts*, **4**(1), 47.

[20] Pratyush, 2010. Analysis and classification of ornaments in north Indian (Hindustani) classical music. Master's thesis, Universitat Pompeu Fabra, Spain.

[21] R. Monelle, 2000. *The Sense of Music: Semiotic Essays*. Princeton, NJ: Princeton University Press.

[22] A. Gabrielsson and P.N. Juslin, 1996. Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychol. Music,* **24**, 68-91.

[23] A. Gabrielsson and E. Lindstrom, 2010. "The role of structure in the musical expression of emotion," in *Handbook of Music and Emotion,* eds P.N. Juslin and J. Sloboda (Oxford, UK: Oxford University Press), pp. 367-400.

[24] D.C. Sergeant and E. Himoni Des, 2016. Gender and music composition: A study of music, and the gendering of meanings. *Frontiers in psychology*, **7**, 411.

The Journal of Acoustical Society of India : Vol. 48, No. 1-2, 2021 (pp. 91-97)

Ragas in Bollywood music — A microscopic view through multifractal cross-correlation method

Shankha Sanyal[1,2*], Archi Banerjee[1,3], Souparno Roy[1,4], Sayan Nag[5], Ranjan Sengupta[1] and Dipak Ghosh[1]

[1]*Sir C.V. Raman Centre for Physics and Music, Jadavpur University, Kolkata, India*
[2]*School of Languages and Linguistics, Jadavpur University, Kolkata, India*
[3]*Rekhi Centre of Excellence for the Science of Happiness, IIT Kharagpur, India*
[4]*Department of Physics, Jadavpur University, Kolkata, India*
[5]*Department of Medical Biophysics, University of Toronto, Canada*
e-mail: archibanerjee7@gmail.com

[Received: 01-11-2020; Revised: 20-12-2020; Accepted: 20-12-2020]

ABSTRACT

Since the start of Indian cinema, a number of films have been made where a particular song is based on a certain *raga*. These songs have been taking a major role in spreading the essence of classical music to the common people, who have no formal exposure to classical music. In this paper, we look to explore what are the particular features of a certain *raga* which make it understandable to common people and enrich the song to a great extent. For this, we chose two common *ragas* of Hindustani classical music, namely "*Bhairav*" and "*Mian ki Malhar*" which are known to have widespread application in popular film music. We have taken 3 minute clips of these two *ragas* from the renderings of two eminent maestros of Hindustani classical music. 3 min clips of ten (10) widely popular songs of Bollywood films were selected for analysis. These were analyzed with the help of a latest non linear analysis technique called Multifractal Detrended Cross correlation Analysis (MFDXA). With this technique, all parts of the Film music and the renderings from the eminent maestros are analyzed to find out a cross correlation coefficient (γ_x) which gives the degree of correlation between these two signals. We hypothesize that the parts which have the highest degree of cross correlation are the parts in which that particular *raga* is established in the song. Also the variation of cross correlation coefficient in the different parts of the two samples gives a measure of the modulation that is executed by the singer. Thus, in nutshell we try to study scientifically the amount of correlation that exists between the *raga* and the same *raga* being utilized in Film music. This will help in generating an automated algorithm through which a naïve listener will relish the flavor of a particular raga in a popular film song. The results are discussed in detail.

1. INTRODUCTION

Raga is said to be the soul of Indian music. Hindi film music, which is commonly referred to as "Bollywood" music, is one of the most popular forms of music in the world today. Almost all Bollywood movies feature several songs that are very popular in India, and quite a large section of these songs are

© 2021 Acoustical Society of India

based on certain *raga*[1]. The classical music system of Indian sub-continent is based on two major concepts - *raga* and *tala*. *Raga* describes the melodic or modal aspect of music and *tala* describes the rhythmic aspect. The rhythmic pattern of any composition in Indian music is described by the term tala, which is composed of cycles of matra's. As the major form of mass entertainment available on a national scale, Hindi cinema plays a prominent and influential role in Indian society. Yet its songs, which represent India's most popular music in the twentieth century, have not been subject to any scientific study. Musicians extensively composed a popular song on the basis of a particular Hindustani raga, the elements of the *raga* used to be predominant in that song which could be identified by professionals. This paper tries to explore a methodology by which one can understand about the essence of a particular raga applied to the popular Bollywood song. Thus it will be possible also for a novice in raga music, to get a flavor of that *raga* just by hearing the popular song. This can work as a process of popularization of raga music. For the study, we chose two common *ragas* namely "*Bhairav*" and "*Mian ki Malhar*" of which one is a time based *raga* conventionally sung in dawn and the other is a seasonal *raga* depicting monsoon. We have taken 3 minute clips from the *'alaap'* portion of vocal renderings of these two *ragas* sung by two eminent maestros of Hindustani music. The *'alaap'* portion of the raga rendition is chosen so as to keep the signal free from tempo variation and also normal accompaniment. Moreover, we get the complete *raga* structure in the *'alaap'* portion. Ten (10) widely popular songs from Bollywood Films are also taken in 3 minute clips[2]. All the music signals were divided into six equal segments of 30 seconds each cut in zero crossing. It is well established that musical signals are generally self-similar. With the possible exception of a few avant-garde compositions, structure and repetition is a general feature of nearly all music[3]. In this context fractal analysis of the signal which reveals the geometry embedded in signal assumes significance.

It is also known that naturally evolving geometries and phenomena are rarely characterized by a single scaling ratio and therefore different parts of a system may be scaling differently. That is, the self similarity pattern is not uniform over the whole system. Such a system is better characterized as 'multifractal'. A multifractal can be loosely thought of as an interwoven set constructed from sub-sets with different local fractal dimensions. Music too, has non-uniform property in its movement. It is therefore necessary to re-investigate the musical structure from the viewpoint of the multifractal theory. Multifractal Detrended Fluctuation Analysis (MFDFA)[4] technique analyzes the musical signal in different scales and gives a multifractal spectral width which is the measure of complexity of the signal. Correlation determines the degree of similarity between two signals. In signal processing, cross-correlation is a measure of similarity of two series as a function of the lag of one relative to the other. A non linear technique called Multifractal Detrended Cross correlation Analysis (MFDXA)[5] is used to analyze the multifractal behaviors in the power-law cross-correlations between two time series data of music signals. With this technique, all segments of the film music and the renderings from the eminent maestros are analyzed to find out a cross correlation coefficient (γ_x) which gives the degree of correlation between these two categories of signals. For uncorrelated data, γ_x has a value 1 and the lower the value of γ_x more correlated is the data[6]. Thus a negative value of γ_x signifies that the two music signals have very high degree of correlation between them. We hypothesize that the songs which show high degree of cross correlation (*i.e.* lower value of γ_x) with the raga clips are the songs in which the essence of that particular raga is present. Also the variation of cross correlation coefficient in the different parts of the two samples gives a measure of the modulation that is executed by the singer. Further, the songs having lower degree of cross correlation with the raga clips are the ones in which the characteristics of that particular raga is absent, may be the features of some other raga or a mixture of *ragas* is present. Thus, in nutshell we propose an automated scientific algorithm with the help of which we can easily check whether the features or the structure of a particular raga is present in popular Hindi film songs.

2. EXPERIMENTAL DETAILS

Ten (10) clips of popular Bollywood songs were taken from[2], each of 3 minute duration. The songs chosen for analysis are detailed in Table 1. A 3 min clip from the *alaap* part of the two *ragas, Bhairav* and *Mian ki Malhar* were taken from the rendition of two eminent vocalists of Hindustani music. The signals

Table 1. Different Bollywood songs chosen for our analysis

Clip No.	Song Name	Singer	Film (Year)
Clip 1	Mohabbat Ki Jhooti Kahani	Lata Mangeshkar	Mughal - E - Azam (1960)
Clip 2	Karo Sab Nichchawr	Asha Bhosle	Ladki Sayadri Ki (1966)
Clip 3	Mohe Bhul Gaye Sanvariya	Lata Mangeshkar	Baiju Bawra (1952)
Clip 4	Naach Mere Mor Zara Naach	Manna Dey	Tere Dwar Khada Bhagwan (1964)
Clip 5	Dil Diya Dard Liya	Md. Rafi	Dil Diya Dard Liya (1966)
Clip 6	Na-Na-Na Barso Badal	Lata Mangeshkar	Prithviraj Chauhan (1959)
Clip 7	Ek Ritu Aaye Ek Ritu	Kishore Kumar	Gautam Govinda (1979)
Clip 8	Duniya Bananewale	Manna Dey	Ziddi (1964)
Clip 9	Insaan Bano	Naushad	Baiju Bawra (1952)
Clip 10	Amma Roti De	Lata Mangeshkar	Sansaar (1952)

are digitized at the rate of 22050 samples / sec 16 bit format. The *alaap* part was considered for analysis because the characteristic features of the entire *raga* is present in this part and that it uses all the notes used in that particular raga and allowed transitions between them with proper distribution over time. Each three minutes signal is divided into six equal segments of 30 seconds each. We measured the cross correlation coefficient for each of the six windows and recorded their variation. The renditions from two different vocalists were chosen to cross check the validation of our hypothesis that the cross correlation coefficient can really be a measure for identification of a particular *raga* on which a song is based on.

3. METHOD OF ANALYSIS

3.1 *Multifractal Detrended Cross Correlation Analysis (MF-DXA)*

We have performed a cross-correlation analysis of correlation between different Bollywood songs and raga clips following the prescription of Zhou[5].

$$x_{avg} = 1/N\sum_{i=1}^{N} x(i) \ \text{ and } \ y_{avg} = 1/N\sum_{i=1}^{N} y(i) \tag{1}$$

Then we compute the profiles of the underlying data series *x(i)* and *y(i)* as

$$X(i) \equiv \left[\sum_{k=1}^{i} x(k) - x_{avg}\right] \text{ for } i = 1.....N \tag{2}$$

$$Y(i) \equiv \left[\sum_{k=1}^{i} x(k) - x_{avg}\right] \text{ for } i = 1.....N \tag{3}$$

The q^{th} order detrended covariance Fq(s) is obtained after averaging over 2Ns bins.

$$F_q(s) = \left\{1/2\mathrm{N_s}\sum_{v=1}^{2NS}[F(s,v)]^{q/2}\right\}^{1/q} \tag{4}$$

where *q* is an index which can take all possible values except zero because in that case the factor *1/q* blows up. The procedure can be repeated by varying the value of *s*. $F_q(s)$ increases with increase in value of *s*. If the series is long range power correlated, then $F_q(s)$ will show power law behavior

$$F_q(s) \sim s^{\lambda(q)}$$

Zhou found that for two time series constructed by binomial measure from p-model, there exists the following relationship[5]:

$$\lambda(q=2) \approx [h_x(q=2) + h_y(q=2)]/2 \tag{5}$$

Podobnik and Stanley have studied this relation when $q = 2$ for monofractal Autoregressive Fractional Moving Average (ARFIMA) signals and EEG time series[7].

In case of two time series generated by using two uncoupled ARFIMA processes, each of both is autocorrelated, but there is no power-law cross correlation with a specific exponent[53]. According to auto-correlation function given by:

$$\lambda(q=2) \approx [h_x(q=2)+h_y(q=2)]/2 \tag{6}$$

The cross-correlation function can be written as

$$C_x(\tau) = \langle [x(i+\tau) - \langle x \rangle][y(i) - \langle y \rangle] \rangle \sim \tau - \gamma_x \tag{7}$$

where γ and γ_x are the auto-correlation and cross-correlation exponents, respectively. Due to the non-stationarities and trends superimposed on the collected data direct calculation of these exponents are usually not recommended rather the reliable method to calculate auto-correlation exponent is the DFA method, namely $\gamma = 2 - 2h$ $(q = 2)$[8]. Recently, Podobnik *et al.*, have demonstrated the relation between cross-correlation exponent, γ_x and scaling exponent $\gamma(q)$ derived from $\gamma_x = 2 - 2\lambda(q = 2)$[7]. For uncorrelated data, γ_x has a value 1 and the lower the value of γ and γ_x more correlated is the data. We know that $h(q) = 0.5$ indicates that the series is an independent random process, and for $h(q) < 0.5$ it is characterized by long-range anti-correlations while for $0.5 < h(q) < 1$, it is featured by long-term correlations. In this case the signal is stationary. The exponent $H(q = 2)$ is equivalent with the well-known Hurst index. A representative figure (Fig. 1) reports the variation of cross correlation exponent $\gamma(q)$ with q for two particular samples (Part 1 for Sample 1 and Sample 2), also the variation of $h(q)$ with q for those two samples obtained from MFDFA technique are also shown in the same figure for comparison.

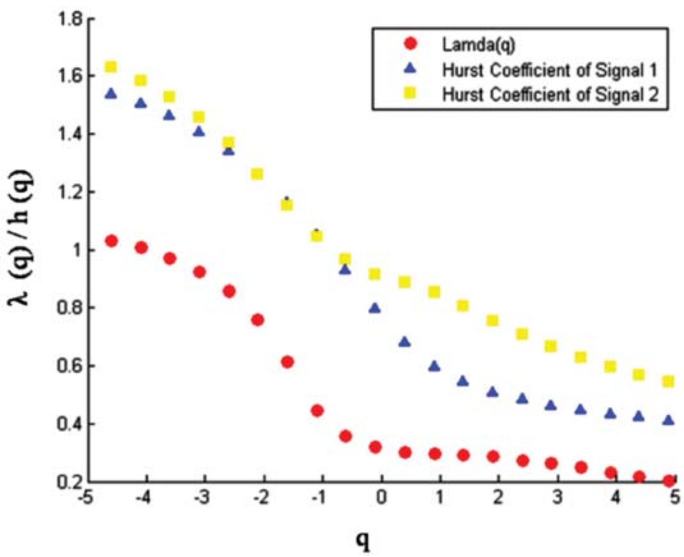
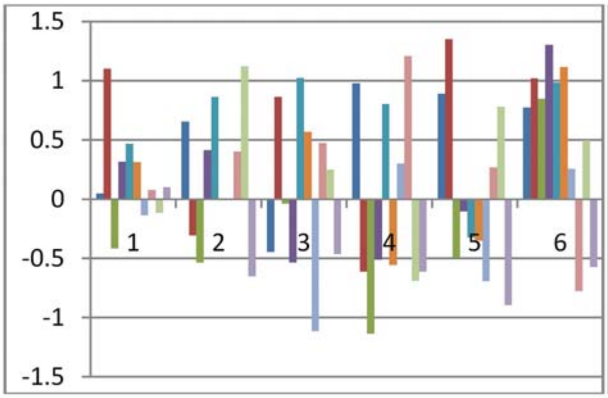


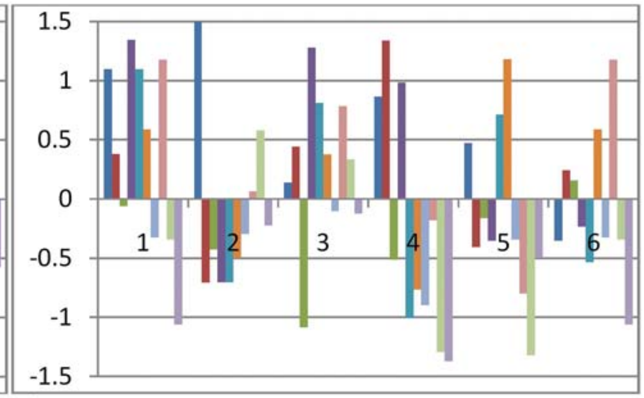
Fig. 1. Variation of $\lambda(q)$ and $h(q)$ for two sound signals

In general, $\gamma(q)$ depends on q, indicating the presence of multifractality. In other words, we want to point out how two sound signals are cross-correlated in various time scales i.e. how much is the essence of a particular raga present in a specific song.

4. RESULTS AND DISCUSSIONS

The cross-correlation coefficient, γ_x was computed for each part of the song and the *raga* clips of the two artistes. Lower the value of γ_x, higher is the degree of cross-correlation between the two signals. A

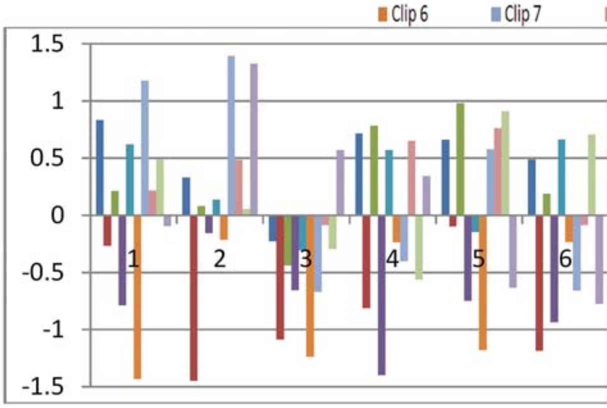


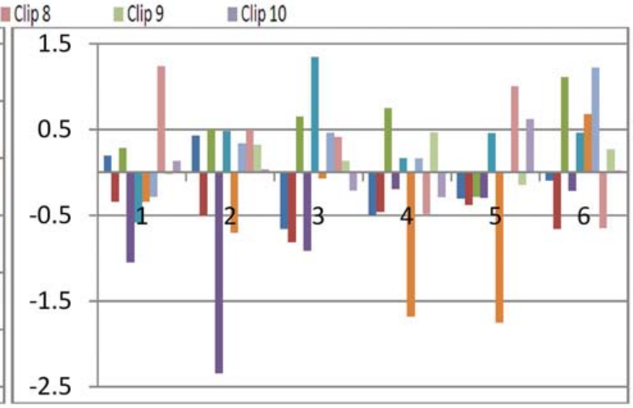


Raga Bhairav

Fig. 2. Variation of γ_x for Artist 1 **Fig. 3.** Variation of γ_x for Artist 2

Clip 1 Clip 2 Clip 3 Clip 4 Clip 5
Clip 6 Clip 7 Clip 8 Clip 9 Clip 10





Raga Mia ki Malhar

Fig. 4. Variation of γ_x for Artist 1 **Fig. 5.** Variation of γ_x for Artist 2

significantly low γ_x would thus signify a greater cue of a particular *raga* being present in that song. The human mind identifies the presence of a *raga* in a song by recognizing certain note structures or transitions that is the characteristic of that *raga.* We want to recreate those features in an automated way, whereby a unique parameter will define the presence of a certain *raga* or a mixture of *ragas* in a musical piece. Fig. 2 and 3 represent the variation of γ_x for the 10 clips and the rendition of *raga Bhairav* by two artistes while Fig. 4-5 represents the same for *Raga Mia ki Malhar*.

The following observations can be drawn from a careful study of the figures :

1. From Figs. 2 and 3, it is seen that Clips 3, 7 and 10 are the ones which in general have negative values of γ_x signifying that they are the ones in which the features of *raga Bhairav* are the strongest. A number of significant dips are also observed in parts of the same song, which can be attributed to very close proximity of the raga and the song. We hypothesize that these are the parts in which a person could relate that a song is based on which particular *raga.*
2. Clip 7 in case of Artist 1 shows a certain amount of ambiguity in belonging to *Raga Bhairav* as is seen from Fig. 3, that the value of γ_x jumps to positive a number of times. But still, since it has considerable low values in other parts, it was considered as based on *raga Bhairav*.
3. It is seen from Figs. 4 and 5 that Clips 2, 4 and 6 show predominant features of *raga Mia ki Malhar* in them. Part 2 of Clip 4, especially has a noteworthy drop in γ_x for Artist 2. This shows that the note structure of *raga Mia ki Malhar* is absolutely followed in this part which results in such a higher degree of cross correlation.

4. Clip 5 shows a unique property of being strongly correlated with both the *ragas* in a number of regions. Thus, we can safely consider that there are elements of different *ragas* woven together in a single piece. If the program is repeated for some more *ragas*, we may have glimpses of more *ragas* present in each of the Clips taken for analysis here.
5. It is also seen from the figures that all the other clips, in some part or the other has some negative values of γ_x, but that may be caused due to spurious matching of note structures of the two clips which has eventually led to higher values of correlation among themselves. It is known that a similar note combination may be used in a number of different *ragas* in Hindustani music, but each *raga* has its own identity which is manifested in the transition between its notes, duration, stress and emphasis on some particular note. All these parameters when taken into consideration will lead to higher degree of cross correlation between two signals which belong to the same *raga*. Hence, we have considered only those clips belonging to a certain *raga* which have consistently shown higher degree of cross-correlation in all the parts.

Next, we have extracted the highest value of cross-correlation from for each combination and put them in Table 2, which will give the readers an estimate of which song is based on which *raga*. The maximum value of γ_x was computed from the six parts for which the program was carried out.

Table 2. Maximum values of γ_x for each *raga*

	Raga Bhairav		***Raga Mia ki Malhar***	
	Artist 1	**Artist 2**	**Artist 1**	**Artist 2**
Clip 1	-0.44	-0.35	-0.22	-0.66
Clip 2	-0.61	-0.7	-1.45	-0.817
Clip 3	-1.13	-1.08	-0.44	-0.28
Clip 4	-0.53	-0.7	-1.39	-1.04
Clip 5	-0.32	-0.53	-0.3	-0.58
Clip 6	-0.34	-0.56	-1.43	-1.75
Clip 7	-1.11	-0.89	-0.67	-0.28
Clip 8	-0.7	-0.17	-0.08	-0.64
Clip 9	-0.11	-0.34	-0.56	-0.14
Clip 10	-0.89	-1.37	-0.77	-0.288

It is seen from Table 2, that whenever the maximum value of γ_x is below -0.8, signature of the presence of a particular *raga* is seen in a musical clip. This is ascertained from our previous observation that these are the clips which showed consistency in maintaining a higher degree of cross correlation in all the parts. This is an interesting observation which can have far reaching consequences when it comes to the study of Hindustani music using non linear analysis. Although, taking the maximum value could be a bit misleading as we are having negative values of γ_x for all the Clips, which implies certain amount of correlation is present always. But again, that can be attributed to the use of similar notes / note-note transition in that particular part where we are getting maximum cross-correlation. Nevertheless, using this technique we have succeeded in obtaining a specific baseline value of -0.80, beyond which any value implies the manifestation of a particular *raga* in a song.

5. CONCLUSION

According to Naushad Ali, a prominent film music composer, ".. classical music has never been the art of masses. It flourished in glamorous courts of Rajas, Maharajas and Nawabs. The common people, who had no access to the great courts were never offered the opportunity of listening to classical music and therefore could not acquire an ear for appreciating it. Hindi film music gave them ample scope to listen to and appreciate classical music[9]". Thus, it has long been said that popularization of classical music can be done through Bollywood music. This work presents a new, interesting data regarding non-linear multifractal analysis of sound signals and the application in the field of *raga* identification from certain Bollywood songs. The following conclusions can be drawn from the study done:

1. We make use of a robust non linear technique MFDXA to quantify the degree of cross correlation between a *raga* clip and a popular Bollywood song. From the cross correlation coefficient, we get a cue for the presence of a particular *raga* in that song.
2. The cross correlation coefficient also gives us a clue about a number of different *ragas* merged together in a certain song which is difficult to be identified by only auditory perception.
3. We have defined a baseline value, beyond which any song will fall in the category of a particular *raga*. Below the stipulated value, the song will have the flavor of more than one *raga*.

In this age of dwindling popularity of classical music, this study could be a boon for popularizing it in the country. If we could generate a one-to-one relation between the success rate of a certain song and the *raga* present in it, the use of that *raga* in more and more songs could be a welcome solution.

6. ACKNOWLEDGEMENTS

Shankha Sanyal acknowledges DST CSRI, Govt of India for providing the funds related to this Major Research Project (DST/CSRI/2018/78 (G)) and the Acoustical Society of America (ASA) for providing the International Students Grant. Archi Banerjee acknowledges the Department of Science and Technology (DST), Govt. of India for providing the DST CSRI Post Doctoral Fellowship (DST/CSRI/PDF-34/2018) to pursue this research work.

7. REFERENCES

[1] Arnold, Alison, 1988. "Popular film song in India: a case of mass-market musical eclecticism." *Popular Music* 7.02, 177-188.

[2] www.chandrakantha.com

[3] Foote Jonathan, 1999. "Visualizing music and audio using self-similarity." *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*. ACM.

[4] J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde and H.E. Stanley, 2002. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, **316**(1), 87-114.

[5] W.X. Zhou, 2008. Multifractal detrended cross-correlation analysis for two nonstationary signals. Physical Review E, **77**(6), 066211.

[6] D. Ghosh, S. Dutta and S. Chakraborty, 2014. Multifractal detrended cross-correlation analysis for epileptic patient in seizure and seizure free status. *Chaos, Solitons & Fractals*, **67**, 1-10.

[7] B. Podobnik and H.E. Stanley, 2008. Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. *Physical review letters*, **100**(8), 084102.

[8] M.S. Movahed and E. Hermanis, 2008. Fractal analysis of river flow fluctuations. *Physica A: Statistical Mechanics and its Applications,* **387**(4), 915-932

[9] Barnouw, Erik and S. Krishnaswamy, 1980. "Indian Film. New York." *Oxford University Press :* pp. 155-165

INFORMATION FOR AUTHORS

ARTICLES

The Journal of Acoustical Society of India (JASI) is a refereed publication published quarterly by the Acoustical Society of India (ASI). JASI includes refereed articles, technical notes, letters-to-the-editor, book review and announcements of general interest to readers.

Articles may be theoretical or experimental in nature. But those which combine theoretical and experimental approaches to solve acoustics problems are particularly welcome. Technical notes, letters-to-the-editor and announcements may also be submitted. Articles must not have been published previously in other engineering or scientific journals. Articles in the following are particularly encouraged: applied acoustics, acoustical materials, active noise & vibration control, bioacoustics, communication acoustics including speech, computational acoustics, electro-acoustics and audio engineering, environmental acoustics, musical acoustics, non-linear acoustics, noise, physical acoustics, physiological and psychological acoustics, quieter technologies, room and building acoustics, structural acoustics and vibration, ultrasonics, underwater acoustics.

Authors whose articles are accepted for publication must transfer copyright of their articles to the ASI. This transfer involves publication only and does not in any way alter the author's traditional right regarding his/her articles.

PREPARATION OF MANUSCRIPTS

All manuscripts are refereed by at least two referees and are reviewed by the Publication Committee (all editors) before acceptance. Manuscripts of articles and technical notes should be submitted for review electronically to the Chief Editor by e-mail or by express mail on a disc. JASI maintains a high standard in the reviewing process and only accept papers of high quality. On acceptance, revised articles of all authors should be submitted to the Chief Editor by e-mail or by express mail.

Text of the manuscript should be double-spaced on A4 size paper, subdivided by main headings-typed in upper and lower case flush centre, with one line of space above and below and sub-headings within a section-typed in upper and lower case understood, flush left, followed by a period. Sub-sub headings should be italic. Articles should be written so that readers in different fields of acoustics can understand them easily. Manuscripts are only published if not normally exceeding twenty double-spaced text pages. If figures and illustrations are included then normally they should be restricted to no more than twelve-fifteen.

The first page of manuscripts should include on separate lines, the title of article, the names, of authors, affiliations and mailing addresses of authors in upper and lowers case. Do not include the author's title, position or degrees. Give an adequate post office address including pin or other postal code and the name of the city. An abstract of not more than 200 words should be included with each article. References should be numbered consecutively throughout the article with the number appearing as a superscript at the end of the sentence unless such placement causes ambiguity. The references should be grouped together, double spaced at the end of the article on a separate page. Footnotes are discouraged. Abbreviations and special terms must be defined if used.

EQUATIONS

Mathematical expressions should be typewritten as completely as possible. Equation should be numbered consecutively throughout the body of the article at the right hand margin in parentheses. Use letters and numbers for any equations in an appendix: Appendix A: (A1, (A2), etc. Equation numbers in the running text should be enclosed in parentheses, i.e., Eq. (1), Eqs. (1a) and (2a). Figures should be referred to as Fig. 1, Fig. 2, etc. Reference to table is in full: Table 1, Table 2, etc. Metric units should be used: the preferred from of metric unit is the System International (SI).

REFERENCES

The order and style of information differs slightly between periodical and book references and between published and unpublished references, depending on the available publication entries. A few examples are shown below.

Periodicals:

[1] S.R. Pride and M.W. Haartsen, 1996. Electroseismic wave properties, *J. Acoust. Soc. Am.*, **100** (3), 1301-1315.
[2] S.-H. Kim and I. Lee, 1996. Aeroelastic analysis of a flexible airfoil with free play non-linearity, *J. Sound Vib.*, **193** (4), 823-846.

Books:

[1] E.S. Skudzryk, 1968. *Simple and Comlex Vibratory Systems*, the Pennsylvania State University Press, London.
[2] E.H. Dowell, 1975. *Aeroelasticity of plates and shells*, Nordhoff, Leyden.

Others:

[1] J.N. Yang and A. Akbarpour, 1987. Technical Report NCEER-87-0007, Instantaneous Optimal Control Law For Tall Buildings Under Seismic Excitations.

SUMISSIONS

All materials from authors should be submitted in electronic form to the JASI Chief Editor: B. Chakraborty, CSIR - National Institute of Oceanography, Dona Paula, Goa-403 004, Tel: +91.832.2450.318, Fax: +91.832.2450.602,(e-mail: bishwajit@nio.org) For the item to be published in a given issue of a journal, the manuscript must reach the Chief Editor at least twelve week before the publication date.

SUMISSION OF ACCEPTED MANUSCRIPT

On acceptance, revised articles should be submitted in electronic form to the JASI Chief Editor (bishwajit@nio.org)