ISSN 0973-3302

THE JOURNAL OF ACOUSTICAL SOCIETY OF INDIA

Volume 50 Number 3-4 July-October 2023

A Quarterly Publication of the ASI

https://acoustics.org.in

The Journal of Acoustical Society of India

Volume 50 Number 3-4 July-October 2023 pp 121-168

The Journal of Acoustical Society of India

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

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

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.

ASI

The Journal of Acoustical Society of India

A quarterly publication of the Acoustical Society of India

Volume 50, Number 3-4, July 2023

ARTICLES

INFORMATION

FOREWORD

The works presented in this third Volume of the three Volume Series of the Special Issue of the Journal of Acoustical Society of India (JASI) look to present new scientific and quantitative research in the domains of music, language and interdisciplinary fields, which were all presented in the 26th International Symposium on Frontiers of Research on Speech and Music (FRSM - 2021), held at IIIT, Pune (virtual mode). The 3rd Volume concludes the Three Volume series of JASI on a high note with chapters focused on latest state of the art methodologies to assess perceptual, psychological and cognitive arousals in response to a auditory stimuli for different categories as well as chapters which deal with traditional Indian ragas and meditation techniques. It may quite be possible that in spite of taking utmost care in choosing the selected and thoroughly revised versions of the manuscript, some inadvertent errors might have cropped in, or there might be certain areas of research that we may have overlooked. If so, the Editors regret the same.

Almost a century ago, the great Indian philosopher and poet Rabindranath Tagore wrote, "Enough study has been done about the musical structure of different ragas. I do not think more study is important in this domain. What is important is that scientists should endeavour to find the reasons why different musical structures evoke different emotions?" In the last few decades, the advent of robust bio-sensors has also provided us with a medium to look into the intricacies and neuronal connections present in the human brain and a way to model the brain. These bio-sensors have enabled scientific endeavours to progress in the path envisaged by Tagore, and a number of studies have shown that listening to music, as well as receiving formal music training, creates an effect of plasticity from the cochlea to the auditory cortex in the human brain. Since the auditory path of musical sounds overlaps functionally with that of speech path, music actually aids in the perception of speech too. Both perceptual and cognitive functions are involved in this process. Music engages a large area of the brain, so music can be used as a supplement in rehabilitation programs and helps the improvement of speech and language skills. For example, studies show that for people with Alzheimer's, music can aid in the process of helping patients access memories that were previously lost. There's also evidence of patients who have suffered brain damage and lost the ability to speak but can still sing when music is played. As elucidated earlier, many excellent studies have shown similar mechanisms between speech and music across many levels. However, a fundamental question, often overlooked, is what makes the brain treat music and speech signals differently and why do humans need two distinct auditory signals. A number of new studies involving advanced neural network theories and computational advances are pointing toward differences in pitch and rhythm as key factors that enable people starting in infancy to distinguish speech from music, as well as how the predictive capabilities of the brain underlie both speech and music perception.

In India, systematic scientific research on music acoustics started after 1920 when Sir C.V. Raman did some pioneering work on Indian string and percussion instruments. After that few Indian scientists, namely Ramakrishna and Kar, did some systematic research on musical acoustics. In 1978, ITC Sangeet Research Academy was formed for the teaching and research on Indian music; it had an equipped Scientific Research Department, probably the only of its kind in India, and academia of traditional gurus to impart training in Gurukul form. They carried out research in various fields of vocal and instrumental music and published their work in various journals. In the year 1990 it was decided at this institute to organise a yearly symposium with an objective to bring researchers in the field of speech and music together. This symposium is held in different parts of the country. The objective of the symposium has always been to encourage interactions from physical, biological, psychological and neuro-sciences in the field of speech and music to provide new ideas and directions. This is the only symposium of its kind, till date, in India in the field. After the closure of the Scientific Research Department of ITC Sangeet Research Academy in 2015, this symposium is continued by Sir C V Raman Centre for Physics and Music, Jadavpur University, in different places in India with the help of other Universities/Institutes. It has completed its 27th year. This concluding volume of the three Volume Series of JASI tries to bring together a number of frontier research works being made in the domains of language, music and their allied applications.

The third volume contains six (6) chapters spanning over the following broad areas:

In the area of speech :

a. Analysis of a Hindustani classical Raga Zeelaf.and studying its resemblances with the melodic modes of Arabic traditional music.

In the area of languages and linguistics :

a. Analysis of the effect of visual noises on viseme recognition on the speakers who utter Malayalam phonemes

b. Perceptual and neural classification of multi-verb constructions - complex predicate constructions (CPC) and serial verb constructions (SVC), constructions which are very much predominant in languages of the South-East Asian family.

In the area of interdisciplinary applications :

a. Study to determine the duration of music training that can bring about improvement in psycho-acoustical tests and also to determine if a relation exists between the performance of music trainees on different psycho-acoustical tests and the rating by their music teacher regarding their musical abilities.

b. Variation of fractal scaling exponents in a song, recitation and reading of a Tagore work with the same lyrical content, pointing to a case of phase transition

c. Comparing the Long Term Effect (LTE) of Sudarshan Kriya (SK), an ancient meditative technique with Music Listening (ML) using EEG analysis.

Last but not least, a few words have to be said about the impact of an Edited Volume like this, which would mainly contribute to the knowledge domain of the nation. As far as the Editors believe, students and young researchers of speech, music, linguistics and allied interdisciplinary areas, particularly in India, and to some extent in other parts of the globe, are in serious need of a reference material that provides comprehensive details about the frontier research trends in the above areas. The Editors hope that this Special Issue of the Journal of Acoustical Society of India (JASI) will be beneficial for them.

Shankha Sanyal, Archi Banerjee,
Sanjeev Sharma and Ranjan Sengupta
—Editors

The Journal of Acoustical Society of India : Vol. 50, No. 3-4, 2023 (pp. 121-127)

A study of the Raga Zeelaf and its relationship with Arabian traditional music

Archi Bhattacharya* and Debesh Kumar Das
CSE Department, Jadavpur University, Kolkata-700 032, India
e-mail: archi.roon@gmail.com

[Received: 17-03-2022; Revised: 01-06-2022; Accepted: 19-06-2022]

ABSTRACT

In this paper we have performed a detailed analysis of a Hindustani classical Raga called *Zeelaf*. It is believed to have been created by the famous Iranian (formerly Persian) poet and musicologist, Hazrat Amir Khusrau. We have thoroughly studied the Raga and have shown that it has resemblances with the melodic modes of Arabic traditional music. The existence of such resemblances certainly gives an idea of the exchange of musical knowledge between the three cultures i.e. Persian, Arabic and Indian and raises a question whether they come from the same root.

1. INTRODUCTION

Amīr Khusro (1253 - 1325), a Sufi musician, great Persian poet and learned scholar from India is sometimes known as "the parrot of India"[1-2]. Son of a Turkish officer in the service of the Sultan of Delhi, Amir was connected with royal courts[1]. Author of five books of poems, composer of some 20000 couplets Amir also composed some famous prose works[1]. As stated in[2], Amir was the first writer to bridge the gulf between two powerful cultures, the Hindu and Muslim. Among his several works, one of his great creations is popular Indian raga Zeelaf, a rarely heard morning raga.

In our earlier work[3], we have shown that there exist some similarities between Hindustani classical music and Arabic traditional music. In this paper, we work on a particular Indian Raga named Zeelaf created by Amir Khusro. We show that Raga Zeelaf has resemblances with Arabic traditional music modes.

2. PRELIMINARIES AND BACKGROUND

In this section we discuss some preliminary requirements and concepts.

2.1 Characterization of music

Any music is characterized by its notes which are linked with different scales. Table 1 shows these dif-ferent scales indicated by the frequencies where each frequency corresponds to a scale and can be the starting note of the music[4]. The frequencies of the other notes are some multiples of the frequency of the starting note. These multiples are given by the following mathematical formula $f_n = f_0 * (a)^n$ where f_0 = the frequency of one fixed note which is pre-defined, f_n = the frequency of the note n half steps away from the fixed note (n may be positive or negative), a = $(2)^{1/12}$ = the twelfth root of 2[4].

Table 1. Notes versus frequencies in Music[4]

Note	Frequency (Hz)	Note	Frequency (Hz)	Note	Frequency (Hz)
C_0	16.35	C_3	130.81	C_6	1046.5
$C^{\#}_0/D^b_0$	17.32	$C^{\#}_3/D^b_3$	138.59	$C^{\#}_6/D^b_6$	1108.73
D_0	18.35	D_3	146.83	D_6	1174.66
$D^{\#}_0/E^b_0$	19.45	$D^{\#}_3/E^b_3$	155.56	$D^{\#}_6/E^b_6$	1244.51
E_0	20.6	E_3	164.81	E_6	1318.51
F_0	21.83	F_3	174.61	F_6	1396.91
$F^{\#}_0/G^b_0$	23.12	$F^{\#}_3/G^b_3$	185	$F^{\#}_6/G^b_6$	1479.98
G_0	24.5	G_3	196	G_6	1567.98
$G^{\#}_0/A^b_0$	25.96	$G^{\#}_3/A^b_3$	207.65	$G^{\#}_6/A^b_6$	1661.22
A_0	27.5	A_3	220	A_6	1760
$A^{\#}_0/B^b_0$	29.14	$A^{\#}_3/B^b_3$	233.08	$A^{\#}_6/B^b_6$	1864.66
B_0	30.87	B_3	246.94	B_6	1975.53
C_1	32.7	C_4	261.63	C_7	2093
$C^{\#}_1/D^b_1$	34.65	$C^{\#}_4/D^b_4$	277.18	$C^{\#}_7/D^b_7$	2217.46
D_1	36.71	D_4	293.66	D_7	2349.32
$D^{\#}_1/E^b_1$	38.89	$D^{\#}_4/E^b_4$	311.13	$D^{\#}_7/E^b_7$	2489.02
E_1	41.2	E_4	329.63	E_7	2637.02
F_1	43.65	F_4	349.23	F_7	2793.83
$F^{\#}_1/G^b_1$	46.25	$F^{\#}_4/G^b_4$	369.99	$F^{\#}_7/G^b_7$	2959.96
G_1	49	G_4	392	G_7	3135.96
$G^{\#}_1/A^b_1$	51.91	$G^{\#}_4/A^b_4$	415.3	$G^{\#}_7/A^b_7$	3322.44
A_1	55	A_4	440	A_7	3520
$A^{\#}_1/B^b_1$	58.27	$A^{\#}_4/B^b_4$	466.16	$A^{\#}_7/B^b_7$	3729.31
B_1	61.74	B_4	493.88	B_7	3951.07
C_2	65.41	C_5	523.25	C_8	4186.01
$C^{\#}_2/D^b_2$	69.3	$C^{\#}_5/D^b_5$	554.37	$C^{\#}_8/D^b_8$	4434.92
D_2	73.42	D_5	587.33	D_8	4698.63
$D^{\#}_2/E^b_2$	77.78	$D^{\#}_5/E^b_5$	622.25	$D^{\#}_8/E^b_8$	4978.03
E_2	82.41	E_5	659.25	E_8	5274.04
F_2	87.31	F_5	698.46	F_8	5587.65
$F^{\#}_2/G^b_2$	92.5	$F^{\#}_5/G^b_5$	739.99	$F^{\#}_8/G^b_8$	5919.91
G_2	98	G_5	783.99	G_8	6271.93
$G^{\#}_2/A^b_2$	103.83	$G^{\#}_5/A^b_5$	830.61	$G^{\#}_8/A^b_8$	6644.88
A_2	110	A_5	880	A_8	7040
$A^{\#}_2/B^b_2$	116.54	$A^{\#}_5/B^b_5$	932.33	$A^{\#}_8/B^b_8$	7458.62
B_2	123.47	B_5	987.77	B_8	7902.13

Hindustani Classical music is characterised by seven main musical notes ('pure' swaras) called the 'saptak' *viz*. Shadja(Sa), Rishabh(Re), Gandhar(Ga), Madhyam(Ma), Pancham(Pa), Dhaivat(Dha) and Nishad(Ni) along with five intermediate notes known as altered notes or 'vikrit swaras'. They are termed as 'Komal' or 'Tivra' viz. Komal Re, Komal Ga, Tivra Ma, Komal Dha, Komal Ni. Thus, there are a total of

12 notes in an octave. A combination of five or more notes upon which a particular melody is based is called a 'Raga' that has its unique ascending and descending structure of the notes and can also be characterized by the melodic patterns of the musical notes. Any classical recital always pertains to a particular raga.

According to the renowned musicologist Pandit Vishnu Narayan Bhatkhande, there are 10 mother ra-gas in Hindustani music (called 'Thaats'[5]) as follows: (1) Asavari (2) Bhairav (3) Bhairavi (4) Bilawal (5) Kafi (6) Kalyan (7) Khamaj (8) Marwa (9) Purvi (10) Todi. All the ragas are derivatives of these 10 Thaats.

2.2 The Raga Zeelaf

The Raga Zeelaf is a rarely heard morning raga created by Amīr Khusrow Dehlavī. It exhibits two forms - one belonging to the Bhairav Thaat and the other belonging to the Asavari Thaat as defined by Bhatkhande's theory. The two forms are described below :

1. ***Zeelaf of Bhairav Thaat :*** It comprises of five dominant notes and two weaker notes (having limited usage). The dominant notes are - Sa, Ga, Ma, Pa and Dha(Komal). One of the characteristic phrases include [Sa Ga Ma Pa Dha(Komal) Pa Ma Ga Ma Sa]. The weaker notes are Komal Re and Ni and are used skilfully to enhance the Bhairav nature of the Raga.
2. ***Zeelaf of Asavari Thaat :*** This uses a total of eight notes *viz*. Sa, Re, Ga (Komal), Ma, Pa, Dha(Komal), Dha, Ni(Komal). One of the characteristic phrases include [Sa Re Ma Pa Dha Ni(Komal) Dha(Komal) Pa Ma Pa Ga(Komal) Re Sa].

2.3 Arabic Traditional Music

The system of melodic modes used in Arabic traditional music is referred to as Maqam. A Maqam is heptatonic, *i.e.*, characterized by 7 musical notes. There are seventy two heptatonic tone rows of Maqamat[6]. Unlike Ragas, most of the Maqams generally start with particular scales and if the scales are different, the Maqams also become different even if the sets of the notes may remain same. Moreover, Maqam scales in traditional Arabic music may be microtonal where the scale is divided into 24 equal quarter tones, and a quarter tone equals half a semitone in a 12 tone equal-tempered scale. Some Maqams include microtonal variations such that tones, half tones and quarter tones are present in its underlying scale.

3. PROPOSED METHOD TO COMPARE RAGA ZEELAF WITH ARABIC MAQAMS

We first perform a signal analysis of Raga Zeelaf following a method almost similar to that followed in[7]. Then the same is done for Maqam. The results of signal analysis of Raga Zeelaf and maqams are then matched.

3.1 Signal Analysis of the Raga Zeelaf

3.1.1 Sample collection : The alap portion (of 1 minute duration) of Raga Zeelaf (both thaats) are played on Harmonium with the Tanpura(drone) playing in the background and recorded using Audacity, an open source software tool working as multi track audio editor and recorder.

3.1.2 F_0 estimation : The audio sample is imported by a software called WaveSurfer, an open source tool for sound visu-alization and manipulation. Then, using the software, the pitch contour is computed.

3.1.3 Steady State detection : The states of the pitch contour which are considerably steady for a minimum of 60 ms are taken and are identified to be the notes which are played in the audio sample. Other regions are discarded.

3.1.4 Tonic extraction : The portion of the audio where the drone is only played is identified and separated into another audio file. Next, a cepstrum i.e. the Fast Fourier Transform of the log of the magnitude spectrum of that portion is computed by applying a Hann window function on the signal[8]. An

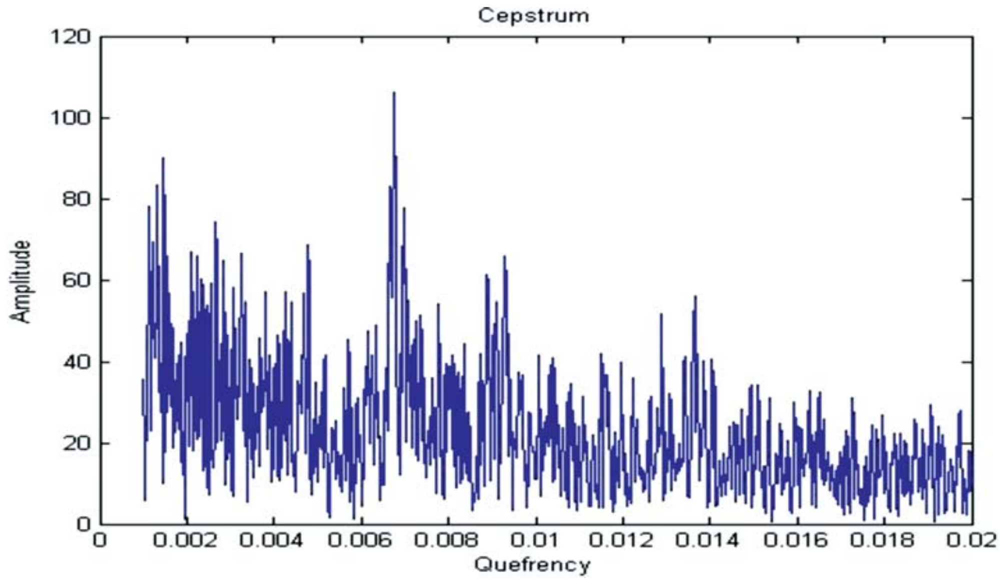


Fig. 1. Cepstrum of the drone portion fft = fft(windowed signal, N), Cepstrum = fft(log(fft)), where N = No. of samples.

example of the cepstrum of the drone portion is shown in Figure 1. Then the frequency corresponding to its highest peak is identified to be the fundamental frequency of the tonic.

3.1.5 *Mapping frequencies into universal notations :* The frequencies obtained from sub section 3.1.3 are mapped into frequencies given in Table 1. The difference between these two sets of frequencies are calculated, and if they are close enough (say, with < 2% difference), then the mapping is done. The notes corresponding to the frequencies in Table 1 where the minimum distances occur for each steady state frequency are entered into an array A and others are discarded. Let the length of the array A be n.

The frequency of the tonic note of the Raga obtained in sub section 3.1.4 is then mapped into universal notation using Table 1.

3.1.6 *Note arrangements & octave adjustments :* The mapped tonic note of the Raga obtained from step *e* of sub section 3.1 is identified in array A. Let this identified note be in the position t in array A ($t < n$).

A new array B of length n is generated from A using the following algorithm

for i = t to n   B[i-t+1] ← A[i]

for i = 1 to t-1 B[i+n-t+1] ← A[i]

In the set of notes obtained in B, a note may appear in more than one octave. In order to get a unique set of notes, only one octave is considered and others are discarded.

3.1.7 *Hindustani note transcription :* The first note of the array B is mapped to 'Sa'. Then the next 11 consecutive notes are found out from Table 1. Based on the distance of the subsequent notes of Table 1 with that of the first note of array B, the other notes are transcribed accordingly.

3.2 Signal Analysis of Maqams

3.2.1 *Sample collection :* Audio samples consisting of the ascending [descending] structures of 10 widely performed maqams are collected. The 10 maqams taken into consideration are: (1) Bayati (2) Hijaz (3) Hijaz Kar (4) Kurd (5) Nahawand (6) Nahawand Kurd (7) Shadd 'Araban (8) Rast (9) Ajam (10) Jiharkah.

3.2.2 F_0 estimation and Steady State detection : The same as described in subsections 3.1.2 and 3.1.3 respectively.

3.2.3 Tonic extraction : Here the tonic is extracted by separating the first note of the audio signal into another audio file. Next, a cepstrum i.e. the Fast Fourier Transform of the log of the magnitude spectrum of that portion is computed using Hann window function. Then the frequency corresponding to its highest peak is identified to be the fundamental frequency of the tonic.

3.2.4 Mapping frequencies into universal notations, Note arrangements & octave adjustments, and Hindustani note transcription : The same as described in subsections 3.1.5, 3.1.6 and 3.1.7 respectively.

3.3 Matching of Maqams with Raga Zeelaf

The transcribed notes of Maqams are matched with the notes data set of Raga Zeelaf.

4. EXPERIMENTAL RESULTS

The Tables 2a and 2b depict the results obtained after performing the analysis of the raga Zeelaf for two forms of the Raga. The first column represents the frequencies of steady state obtained from the audio signals those are mapped with Table 1 as shown in column 2. The difference between steady state frequencies and mapped frequencies are found to be less than 2% as shown in column 3 and corresponding

Table 2a. Signal Analysis of Zeelaf_asavari thaat.

Steady State frequencies	Mapping in Table 1	% Difference	Notes in Table 1 (array A)	Re-ordering of notes (array B)	After Octave adjustments	Transcribed notes
131.0821571	130.81	0.21	C_3	D_3	D	Sa
146.1383832	146.83	0.47	D_3	E_3	E	R2
164.1929102	164.81	0.37	E_3	F_3	F	G1
174.0464423	174.61	0.32	F_3	G_3	G	M1
195.3565251	196	0.33	G_3	A_3	A	Pa
220.0392781	220	0.02	A_3	$A^{\#}_3$	$A^{\#}$	D1
232.7692308	233.08	0.13	$A^{\#}_3$	B_3	B	D2
247.3836121	246.94	0.18	B_3	C_4	C	N1
259.3531116	261.63	0.87	C_4	C_3		

Table 2b. Signal Analysis of Zeelaf_bhairav thaat.

Steady State frequencies	Mapping in Table 1	% Difference	Notes in Table 1 (array A)	Re-ordering of notes (array B)	After Octave adjustments	Transcribed notes
123.3147935	123.47	0.13	B_2	$D^{\#}_3$	$D^{\#}$	Sa
146.4808856	146.83	0.24	D_3	E_3	E	R1
155.6103268	155.56	0.03	$D^{\#}_3$	G_3	G	G2
164.7957249	164.81	0.01	E_3	$G^{\#}_3$	$G^{\#}$	M1
195.9608575	196	0.02	G_3	$A^{\#}_3$	$A^{\#}$	Pa
205.1767731	207.65	1.19	$G^{\#}_3$	B_3	B	D1
231.9766098	233.08	0.47	$A^{\#}_3$	D_4	D	N2
246.8907241	246.94	0.02	B_3	$D^{\#}_4$		
293.9973668	293.66	0.11	D_4	B_2		
311.3971641	311.13	0.09	$D^{\#}_4$	D_3		

Table 3. Comparison of Maqam with Zeelaf.

Maqams name	Steady State frequencies	Mapping in Table 1	% Diffe-rence	Notes in Table 1 (array A)	Re-ordering of notes (array B)	After Octave adjustments	Trans-cribed notes	Resem-blance with
Naha-wand Kurd	130.788669	130.81	0.02	C_3	C_3	C	Sa	Zee-laf_asavari thaat (par-tially)
	145.986751	146.83	0.57	D_3	D_3	D	R2	
	155.079239	155.56	0.31	$D^{\#}_3$	$D^{\#}_3$	$D^{\#}$	G1	
	174.064733	174.61	0.31	F_3	F_3	F	M1	
	196.490969	196	0.25	G_3	G_3	G	Pa	
	207.364516	207.65	0.14	$G^{\#}_3$	$G^{\#}_3$	$G^{\#}$	D1	
	233.611374	233.08	0.23	$A^{\#}_3$	$A^{\#}_3$	$A^{\#}$	N1	
Naha-wand	130.395745	130.81	0.32	C_3	C_4	C	Sa	Zee-laf_asavari thaat (par-tially)
	196.544978	196	0.28	G_3	D_4	D	R2	
	208.30677	207.65	0.32	$G^{\#}_3$	$D^{\#}_4$	$D^{\#}$	G1	
	232.955882	233.08	0.05	$A^{\#}_3$	F_4	F	M1	
	246.598827	246.94	0.14	B_3	C_3	G	Pa	
	260.634725	261.63	0.38	C_4	G_3	$G^{\#}$	D1	
	292.728636	293.66	0.32	D_4	$G^{\#}_3$	$A^{\#}$	N1	
	309.370642	311.13	0.57	$D^{\#}_4$	$A^{\#}_3$	B	N2	
	348.850672	349.23	0.11	F_4	B_3			
Shadd 'Araban	184.331019	185	0.36	$F^{\#}_3$	G_3	G	Sa	Zee-laf_bhairav thaat
	196.616008	196	0.31	G_3	$G^{\#}_3$	$G^{\#}$	R1	
	208.732867	207.65	0.52	$G^{\#}_3$	B_3	B	G2	
	245.474941	246.94	0.59	B_3	C_4	C	M1	
	259.49292	261.63	0.82	C_4	D_4	D	Pa	
	293.295309	293.66	0.12	D_4	$D^{\#}_4$	$D^{\#}$	D1	
	310.255816	311.13	0.28	$D^{\#}_4$	$F^{\#}_3$	$F^{\#}$	N2	
Hijaz Kar	131.373729	130.81	0.43	C_3	C_3	C	Sa	Zee-laf_bhairav thaat
	138.472688	138.59	0.08	$C^{\#}_3$	$C^{\#}_3$	$C^{\#}$	R1	
	162.330912	164.81	1.5	E_3	E_3	E	G2	
	172.937266	174.61	0.96	F_3	F_3	F	M1	
	195.517553	196	0.25	G_3	G_3	G	Pa	
	208.021654	207.65	0.18	$G^{\#}_3$	$G^{\#}_3$	$G^{\#}$	D1	
	243.70452	246.94	1.31	B_3	B_3	B	N2	
	260.428663	261.63	0.46	C_4	C_4			
Rast	130.694305	130.81	0.09	C_3	C_4	C	Sa	Zee laf_asavari thaat (par-tially)
	155.585337	155.56	0.02	$D^{\#}_3/E^{b}_3$	D_4	D	R2	
	174.309364	174.61	0.17	F_3	F_4	$D^{\#}$	G1	
	196.514018	196	0.26	G_3	C_3	F	M1	
	220.452483	220	0.21	A_3	$D^{\#}_3/E^{b}_3$	G	Pa	
	233.013231	233.08	0.03	$A^{\#}_3/B^{b}_3$	F_3	A	D2	
	261.163233	261.63	0.18	C_4	G_3	$A^{\#}$	N1	
	293.45006	293.66	0.07	D_4	A_3			
	347.534233	349.23	0.49	F_4	$A^{\#}_3/B^{b}_3$			

Here, R1=Komal Re, R2=Re, G1=Komal Ga, G2=Ga, M1=Ma, M2=Tivra Ma, D1=Komal Dha, D2=Dha, N1=Komal Ni, N2=Ni.

notes are extracted from column 1 of Table 1 and reported in column 4 (array A). The shaded cells of column 2 represent the tonics, corresponding to D_3 and $D^{\#}_3$ in column 3 of Tables 2a and 2b respectively. The notes in array A are reordered in column 5 following the algorithm described in subsection 3.1.6. We then make octave adjustment as shown in column 6. The first note being 'Sa', the transcribed notes are stated in last column.

Table 3 depicts the results obtained after analyzing the different maqams. The maqam names are given in column 1. The same procedure as followed in case of Raga Zeelaf is done for every maqam with only difference in the tonic extraction technique as described in subsection 3.2.3. The transcribed notes of different maqams are reported in last column.

The Zeelaf_bhairav thaat are found to be completely matching with maqams Shadd 'Araban and Hijaz Kar. Nahawand Kurd, Nahawand and Rast have partially matched with Zeelaf_asavari thaat.

5. CONCLUSION

We have compared Raga Zeelaf created by Amīr Khusrow with Arabian traditional music modes known as maqams. We have found that there exists quite resemblances between them.

The existence of such resemblances certainly proves the exchange of musical knowledge between the three cultures i.e. Persian, Arabic and Indian. As stated by historians, there exists a lot of similarities in the Indian and Iranian (formerly called as Persian) languages[9]. Our work in this paper, establishes the similarities in music melodies of these two countries as well as the Arab world. The existence of such resemblances certainly raises a question whether they are having the same root. The research on this may clearly throw some light on the study of both history and music. The work also opens up some future plans to make extensive study with more examples and also to study the other kinds of music and find their resemblances with Indian music.

6. REFERENCES

[1] Thedore Pappas, 2006. Ed., *Encyclopaedia Britannica*, Encylopaedia Britanica (India) Pvt. Ltd., New Delhi and Impulse Marketing, New Delhi, **1,** 78.

[2] Suniti Kumar Chatterjee, 2005. Ed., *The Cultural Heritage of India*, The Ramakrishna Mission Institute of Culture, Calcutta, **5,** 491-492.

[3] Archi Bhattacharya and Debesh Kumar Das, 2017. Do Traditional Arabic Music and Hindustani Clas-sical Music have Same Root? *Proceedings of 23rd International Symposium on Frontiers of Research in Speech and Music (FRSM)*, NIT Rourkela.

[4] https://pages.mtu.edu/~suits/notefreqs.html

[5] Vishnu Narayan Bhatkhande, 1954. *Hindustani Sangeet Paddhati,* Kramik Pustak Malika, Laxmi Narayan Garg, Sangeet Karyalaya Hathras, p. 484.

[6] Habib Hassan Touma, 1996. *The Music of the Arabs*, Amadeus Press, Portland, Oregon, USA, p. 262.

[7] Archi Bhattacharya and Debesh Kumar Das, 2016. Recognition of Ragas of Hindustani Music Played on Harmonium, *Proceedings of 22nd International Symposium on Frontiers of Research in Speech and Music (FRSM),* Baripada, Orissa, India, November.

[8] R. B. Blackman and J. W. Tukey, 1959. *The Measurement of Power Spectra,* Dover Publication Inc., New York, p. 190.

[9] Irfan Habib, 2011. *A People's History of India,* **2.** The Indus Civilization, Tulika Books, New Delhi, p. 111.

The Journal of Acoustical Society of India : Vol. 50, No. 3-4, 2023 (pp. 128-137)

Impact of visual noise on Malayalam viseme recognition

Bibish Kumar K.T.[1], R.K. Sunil Kumar[2], Sunil John[1] and Muraleedharan K. M.[1]
[1]Computer Speech and Intelligence Research Centre, Department of Physics, Government College Madappally, Vadakara, Calicut, Kerala, India
[2]Department of Information Technology, Kannur University, Kerala, India
e-mail: bibishkrishna@gmail.com

[Received: 01-03-2022; Revised: 10-05-2022; Accepted: 02-06-2022]

ABSTRACT

The talker's face significantly aids in recognizing what is spoken, especially in acoustically noisy conditions. The last decade witnessed significant efforts that employed visual cues to improve speech-based applications like the AVSR (audio-visual speech recognition) system's performance. However, the impact of visual noise on speech recognition system performance has not received sufficient attention in the literature. This study analyzes the effect of visual noise on the viseme recognition of 23 speakers who utter Malayalam phonemes. "MOZHI" an audio-video speech database, was used for this purpose. Linguistically, the 50 Malayalam phonemes are mapped into 14 viseme elements. Three different types of visual noise with varying intensities were simulated to represent the corruption likely to occur in a real-world application: blur (out of focus), jitter (camera / speaker movement), and random Gaussian noise (degradation of image quality). The performance of the viseme recognition system has been evaluated in clean and noisy visual conditions by extracting DCT features and applying an SVM classifier. It was found that jitter corruption significantly affected the viseme recognition rate when compared to the two other visual noises.

1. INTRODUCTION

Speech processing is a distinct discipline that covers a broad area by incorporating various technologies and applications that enable humans to interact seamlessly with intelligent systems. One of the most challenging features, especially for researchers, in the speech processing realm is its multidisciplinary nature, which necessitates knowledge and expertise from various disciplines. Research and development of ASR (automatic speech recognition) systems began in the 1950s at Bell Labs, with simple digit recognition systems (Aida-Zade, Xocayev, and Rustamov 2017). Since then, recognition tasks have become more complex, moving from speaker-dependent isolated word recognition to speaker-independent continuous speech recognition, and from an extensive vocabulary to spontaneous speech recognition.

The existing audio-only speech-based application system is only successful in relatively controlled surroundings or under calm conditions. Any intense background noise has a negative impact on the performance of such a system. Humans, on the other hand, can typically compensate for such ambiguity by incorporating additional speech information sources, such as the speaker's facial appearance. To recognize speech, humans integrate both visual and audio cues wisely. This effect is known as the *"McGurk effect"* (Bear and Harvey 2016). Lips play a vital role in overcoming the limitations associated with audio-only speech perception. This idea of using visual cues inspired many researchers to devote their efforts to developing intelligent systems using audio and visual speech information, namely audio-visual speech recognition (AVSR).

Following Petajan's first attempt at a visual speech recognition system in 1984 (Potamianos *et al.* 2017), a wide range of AVSR systems have been produced, and the results confirmed that AVSR systems have a better recognition rate than audio-only speech recognition systems, especially in noisy environments. Visual speech mainly conveys information about the point of articulation, and this can easily be confused when the audio modality is noisy.

However, questions like what happens if the visual speech signals are corrupted and how much it affects the performance of an AVSR system are still inadequately explored. In this work, three visual noises of varying strengths are added to clean visual speech, and their effects are explored using the visual-only speech recognition (VSR) system. The visual noises for the study comprise blur (out of focus), jitter (camera / speaker movement), and random Gaussian noise (degradation of image quality). These noises are chosen as they are relevant to real-time visual distortion.

This work focuses on the Malayalam language, which is a Dravidian language spoken in India and serves as the official language of Kerala. A Malayalam audio-visual speech database named "MOZHI" is used to carry out this work (Bibish Kumar *et al.*, 2021). However, processing the entire video signal is computationally expensive in any real-time speech-processing application. Therefore, knowledge about visually separable basic units in a language (viseme) is vital in creating any multimodal speech-based applications. A phoneme is the basic sound unit necessary to symbolize all words in that speech (Bear and Harvey 2017). The corresponding language unit for visual speech is termed a viseme. In Malayalam, to analyze visual speech, it is necessary to establish a relationship between a phoneme and its visual equivalent by establishing a viseme set. This work also considers the linguistic classification of 50 Malayalam phonemes into 14 visemes. Most relevant frames (visemes) are identified based on linguistic knowledge and input to the discrete cosine transform (DCT) to extract relevant attributes, resulting in a linguistically involved data-driven approach. In this study, the support vector machine (SVM) classifier is utilized to identify the underlying viseme in both clean and corrupted visual speech.

The paper structure looks as follows: Section 2 discusses the database used in this work. Section 3 provides a brief outline of the linguistic background of the Malayalam language. Section 4 introduces the different visual noises used in the work. Section 5 displays the experimental work, and Section 6 summarizes the outcomes of the study.

2. MOZHI DATABASE

The main criterion for speech database construction is a large, phonetically balanced corpus uttered by many distinct speakers in an uncontrolled environment. In addition, it is necessary to understand the peculiarities of the language of the database and its linguistic background. The need for diversity in resources and massive storage capacity is the main challenge that produces the remarkable difference in statistics between audio-only and audio-visual speech databases. Researchers in this field are focused on developing speech-based applications for European languages, particularly English. However, speech processing in under-resourced languages like Malayalam is still in its early stages, with only a few works focusing on audio speech. Under-resourced language speech processing is hindered by the absence of standard audio-visual speech databases and computational linguistic resources.

Malayalam, a Dravidian language spoken across Kerala, Lakshadweep, and Mahe, is spoken by 38 million people worldwide and has possessed the status of a classical language since 2013. Malayalam is syllabic, with exact correspondence between spoken and written syllables. It has 50 phonemes and 106 allophones (K. T. Bibish Kumar *et al.* 2019). Phonemes are the relatively distinct and fundamental utterances of a language. An allophone is a phonetic variant that is phonetically distinct (Skrelin 1999). The study makes use of "MOZHI," a Malayalam audio-visual speech database, and the utterances are recorded in various environments for various research purposes, making it the first of its kind in the language. This study uses the first category of the database, which consists of the utterances of 18 female speakers and five male speakers who each say 50 phonemes three times. These are recorded in a controlled environment. In the Malayalam language, there are ten vowel phonemes, two diphthongs, and 38 consonant phonemes. Based on articulation, four out of ten vowels are classified as front vowels, two as central vowels, and the remaining four as back vowels. Similarly, the consonant phonemes are categorized as bilabial, labiodental, dental, alveolar, retroflex, palatal, velar, and glottal based on the relative position of the articulators.

The high-quality visual speech signals are recorded at a frame rate of 25 fps and a resolution of 1280 x 720. The audio signal is sampled at 44100 Hz. This camera is used to capture the dynamic variation of the mouth region with audio and video signals wrapped in MP4 format. The approximate length is five minutes for isolated phonemes for each speaker. Each isolated phoneme and word in the audio and video domains are segmented and labeled.

3. PHONEME-TO-VISEME MAPPING

Recent research in visual language has brought about considerable alterations to the definition of viseme. A viseme can be considered in terms of articulatory gestures such as mouth opening, teeth, and tongue vulnerability that must generate different phonemes (Fisher 1968). An equivalent definition used extensively in literature is a viseme for a set of phonemes with a similar visual look (Bear and Harvey 2016). The current description of viseme is a lively visual language unit that describes distinct speech movements of the visual speech articulators. The lips, tongue, teeth, and jaw are the active articulators

Table 1. Linguistic Classification of Vowel and Diphthong Phonemes Visually

Viseme	Viseme Class	Phoneme with IPA	Selected Frame
1	Front, High– Vowel	ഇ /i/, ഈ /i:/	
2	Front, Mid – Vowel	എ /e/, ഏ /e:/	
3	Central, Low – Vowel	അ /a/, ആ /a:/	
4	Back, High – Vowel	ഉ /u/, ഊ /u:/	
5	Back, Mid – Vowel	ഒ /o/, ഓ /o:/	
6	Diphthong 1	ഐ /ai/	
7	Diphthong 2	ഔ /au/	

used in language production (Taylor *et al.* 2012). Analysis of the lips, the most active visible articulator, is crucial in the visual speech analytics framework for recognition and synthesis. Due to the appearance of the lips, many phonemes belong to a single viseme class, thereby creating a bridge between phoneme and viseme, which is termed phoneme-to-viseme mapping or many-to-one mapping.

In Malayalam, the language component of audio speech is linguistically classified according to articulation points and manners, as described in (K. T. Bibish Kumar *et al.* 2019). The visual aspect of speech depends primarily on lip and jaw movements. The visibility of teeth and tongue is also a vital element. While uttering vowel phonemes, the lips are either wide open or projected outward. However, consonant phonemes are produced by touching the active articulator tongue at different places inside the mouth area, whose dynamics are not visible. Accordingly, the extent of the appearance of active articulators is the determining factor for consonant phonemes. This section explores the possibilities of forming a viseme set from linguistic knowledge by consulting linguists. The viseme set for vowels and consonants in Malayalam formed from linguistic understanding is given in Tables 1 and 2, respectively.

Table 2. Linguistic Classification of Consonant Phonemes Visually

Viseme	Viseme Class	Phoneme with IPA	Selected Frame
8	Bilabial-Plosive-voiced and voiceless unaspirated, Nasal	പ്/p/, ബ് /b/, ഭ് /b^h/, മ്/m/	
9	Bilabial-Plosive-voiceless aspirated And Labiodental	ഫ്/p^h/, വ്/v/	
10	Dental	ത്/t/, ഥ്/t^h/, ദ്/d/, ധ്/d^h/, ന്/n□/	
11	Velar Glottal	ക്/k/, ഖ്/kh/, ഗ്/g/, ഘ്/gh/, ങ്/ŋ/, ഹ് /h/	
12	Alveolar	□്/□/, ന്/n/, സ്/s/, ര്/r/, റ്/□/, ല്/l/	
13	Retroflex	ട്/□/, ഠ്/□h/, ഡ്/□/, ഢ്/□h/, ണ്/□/, ഷ്/□/, ള്/□/, ഴ്/□/	
14	Palatal	ച്/c/, ഛ്/ch/, ജ്/□/, ഝ്/□h/, ഞ്/□/, ശ്/□ /, യ്/y/	

Based on the linguistic mapping, frames of similar visual appearance are selected manually for each speaker. As the visual lengths and appearances of vowel phonemes and consonant phonemes differ significantly, selected frames alone may cause sparse classification from a computational point of view.

Fig. 1. Visual representation of Consonant Phoneme ച്/c/ (Middle frame as Selected frame).

Additionally, owing to computational complexity, it is not advisable to consider every frame in a phoneme as visual speech. The time evolution of the chosen frame is used to solve this problem. Accordingly, a viseme is represented by the selected frame and two preceding and following frames (5 frames), as in Figure 1.

4. VISUAL NOISE

In signal processing, noise is considered as a signal with various frequency components at varying strengths that can distort the nature of the original signal during the recording, processing and transmission of the signals. In the real world, various noises distort speech signals, reducing their perceptual quality and intelligibility. In this work, three real-time visual noises of varying strengths were added to clean visual speech, and then DCT coefficients were extracted to examine the performance of the VSR systems in these conditions.

Blurring simulates a real-life scenario in which the camera loses focus, or the speaker tilts his head back and forth while recording. Depending on the strength of these causes, the region of the video frame encompassing the speaker's mouth will be opaque. Gaussian filters with five different standard deviation values (2, 4, 6, 8, and 10) were used to distort clean visual speech. Figure 2 depicts the effect of blurring on clean visual speech at various intensities.

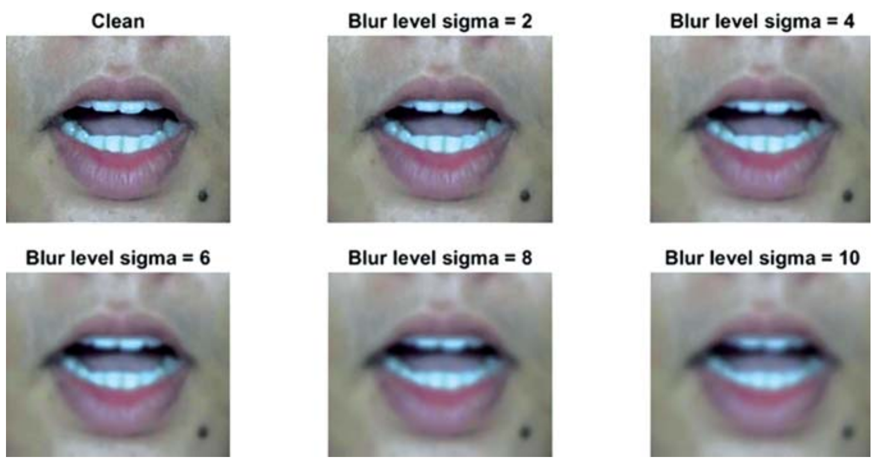


Fig. 2. Illustration of Impact of Blur noise at various intensities.

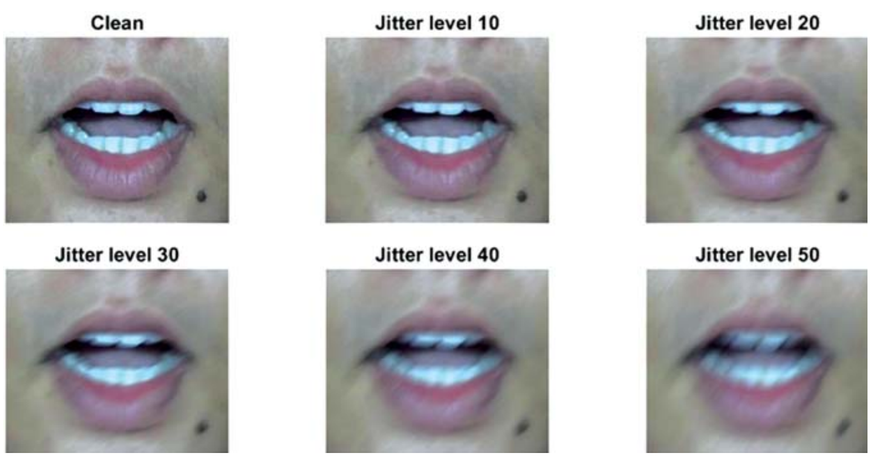


Fig. 3. Illustration of Impact of Jitter noise at various intensities.

In a practical sense, jittery noise represents either camera shakes or the speaker's head movement. Jitter noise is generated by rotating a camera / frame in a counter clockwise manner while moving it linearly in terms of pixels. The visual speech that results provides the sensation that the speaker's mouth is moving in each frame. In clean visual speech, five jitter levels are used: 10, 20, 30, 40, and 50. Jitter level 10 equates to 10-pixel horizontal motion and 10-degree rotation. Figure 3 shows the effect of jitter noise on clean visual speech at various intensities.

Gaussian noise is used to degrade the quality of an image. Gaussian noise harms the ability of humans to distinguish the lip region from its surroundings. In this study, five levels of Gaussian noise with variances ranging from 0.02 to 0.10 are applied to clean visual speech. Figure 4 shows the influence of Gaussian noise on clean visual speech at different intensities.

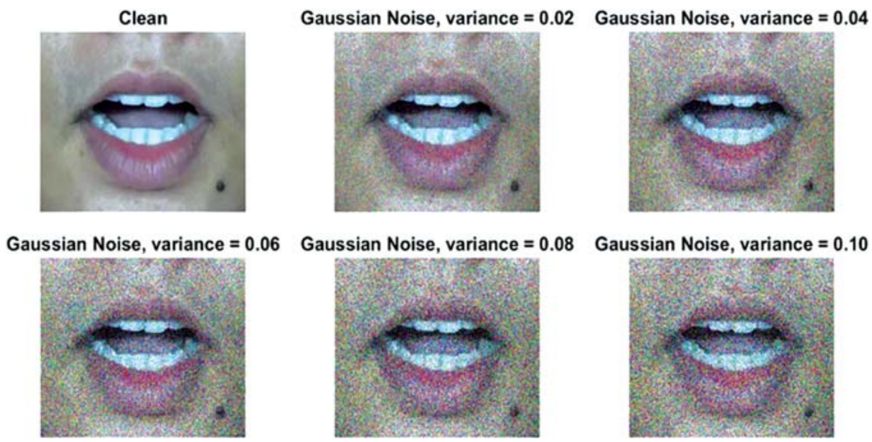


Fig. 4. Illustration of Impact of Gaussian noise at various intensities

5. EXPERIMENTAL RESULTS

Training and testing sets are created after the dataset has been prepared. The training set is inserted into the machine learning algorithm to create a predictive model. The model then predicts the labels in the testing data. The model's performance is then evaluated using precision, recall, F1-score, and accuracy. In this work, audio speech samples were collected from 23 speakers uttering 50 utterances (vowels, diphthongs, and consonant phonemes) that were repeated three times each. Thus, each speaker has a total of 150 utterances / observations. These video speech samples were corrupted with three different noises at five noise levels. Since the viseme for each phoneme is represented by five consecutive frames, later, the visual representation of each phoneme is grouped based on phoneme-to-viseme mapping as in Table 1 and Table 2. Thus, the visual speech sample consists of 750 frames (50 phonemes x 3 repetitions x 5 frames) for each speaker. First 20 discrete cosine transform (DCT) coefficients are extracted from each image to represent the visual speech signals. Each viseme contains an unequal number of phonemes, resulting in an unbalanced dataset with a minimum and maximum of 69 (1 phoneme x 23 speakers x 3 repetitions for viseme 6 and 7) and 552 observations (for viseme 13), respectively.

For evaluating the performance of the SVM classifier: accuracy, precision, recall (sensitivity), and F1-score were used and are defined below:

$$\text{Accuracy} = TP+TN / (TP+TN+FP+FN) \tag{1}$$

$$\text{Precision} = TP / (TP+FP) \tag{2}$$

$$\text{Recall} = TP / (TP+FN) \tag{3}$$

$$\text{F1-score} = 2*(\text{Recall} * \text{Precision}) \ / \ (\text{Recall} + \text{Precision}) \tag{4}$$

where TP is True Positive, FN is False Negative, FP is False Positive, and TN is True Negative.

The SVM classifier is used in this study to recognize the viseme from corrupted visual noise. The SVM classifier's main feature is the selection of optimal hyperspace parameters such as margin width from penalty parameter C and margin shape (gamma) which are determined experimentally using grid search (Aida-Zade, Xocayev, and Rustamov 2017). The estimated parameter values are $C = 10^2$ and gamma $= 10^{-2}$, along with the Gaussian radial basic kernel. The SVMs are trained with 60% of the dataset and classification results in terms of precision, recall, F1-score, and accuracy are shown in Figure 5.

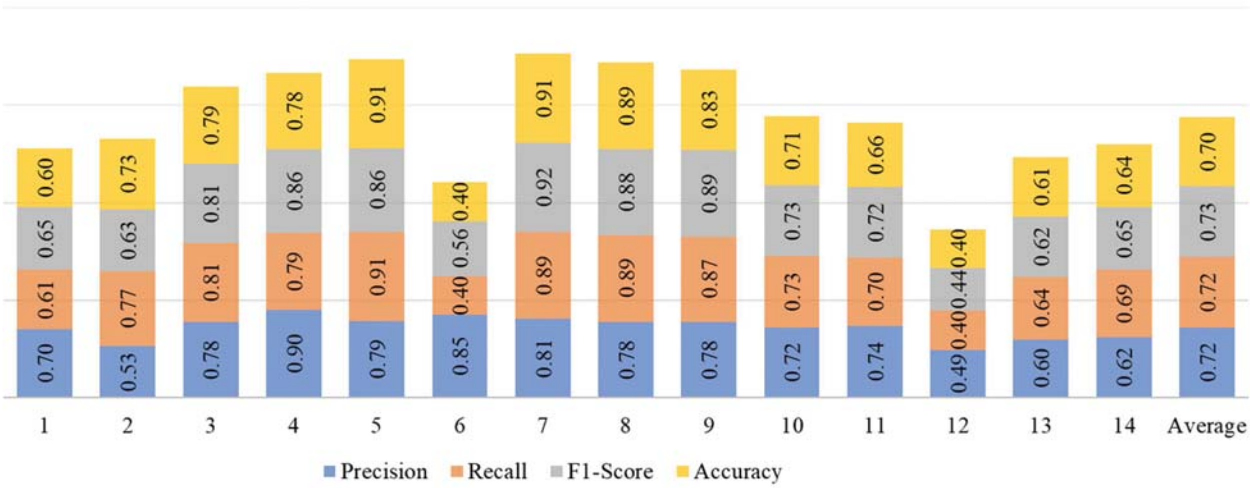


Fig. 5. Viseme Recognition from Clean Visual Speech using SVM Classifier.

We adopt a modified version of SVM that addresses the issues of optimum hyperparameters, overfitting, and unbalanced datasets by using nested stratified 5-fold cross-validation (K. T. Kumar 2021) such that each observation serves as either training data, validation data, or test data. Nested cross-validation overcomes the issue of overfitting as in cross-validation, where the same test set is used for both model selection and estimation, resulting in a biased outcome. Stratified 5-fold cross-validation rearranges the data in each class, thereby ensuring that each fold is a reliable representative of the whole. The performance of the proposed VSR system in a clean visual environment based on a modified SVM classifier is shown in Figure 6.

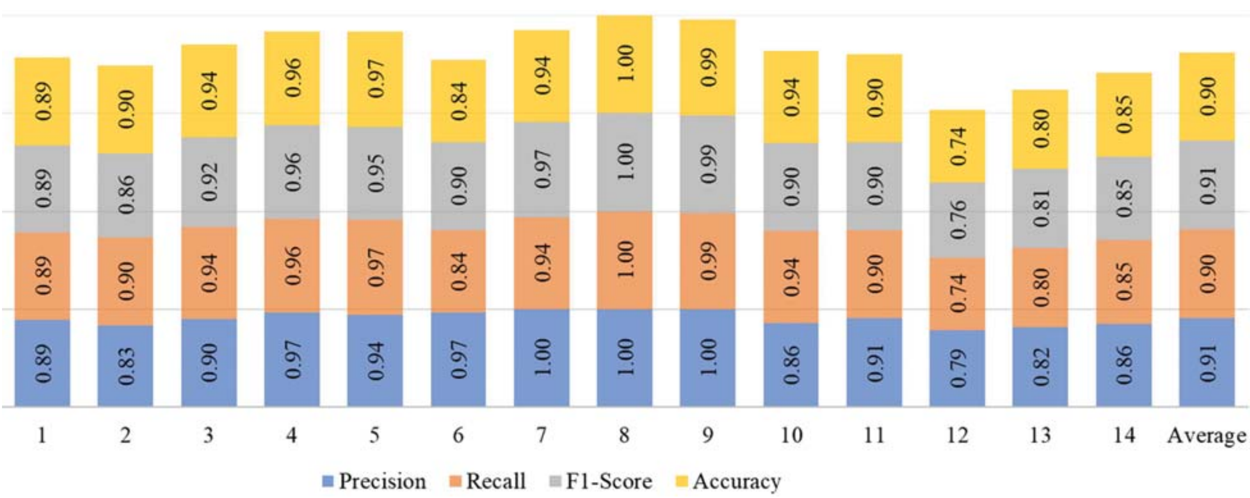


Fig. 6. Viseme Recognition from Clean Visual Speech using Modified SVM Classifier.

Table 3. Performance of Viseme Recognition System in Visual Noise

Noise Type		Viseme Recognition Average Accuracy (%)
Clean		90
Blurring (Standard Deviation)	2	80
	4	69
	6	64
	8	59
	10	56
Jitter	{10, 10°}	84
	{20, 20°}	76
	{30, 30°}	46
	{40, 40°}	23
	{50, 50°}	10
Random Gaussian Noise (Variance)	0.02	78
	0.04	76
	0.06	68
	0.08	60
	0.1	51

The experimental results show that viseme recognition using a modified SVM classifier gives a better result, with an average of 90% in all metrics. Table 3 and Figure 7 display the performance of the proposed system in corrupted visual speech.

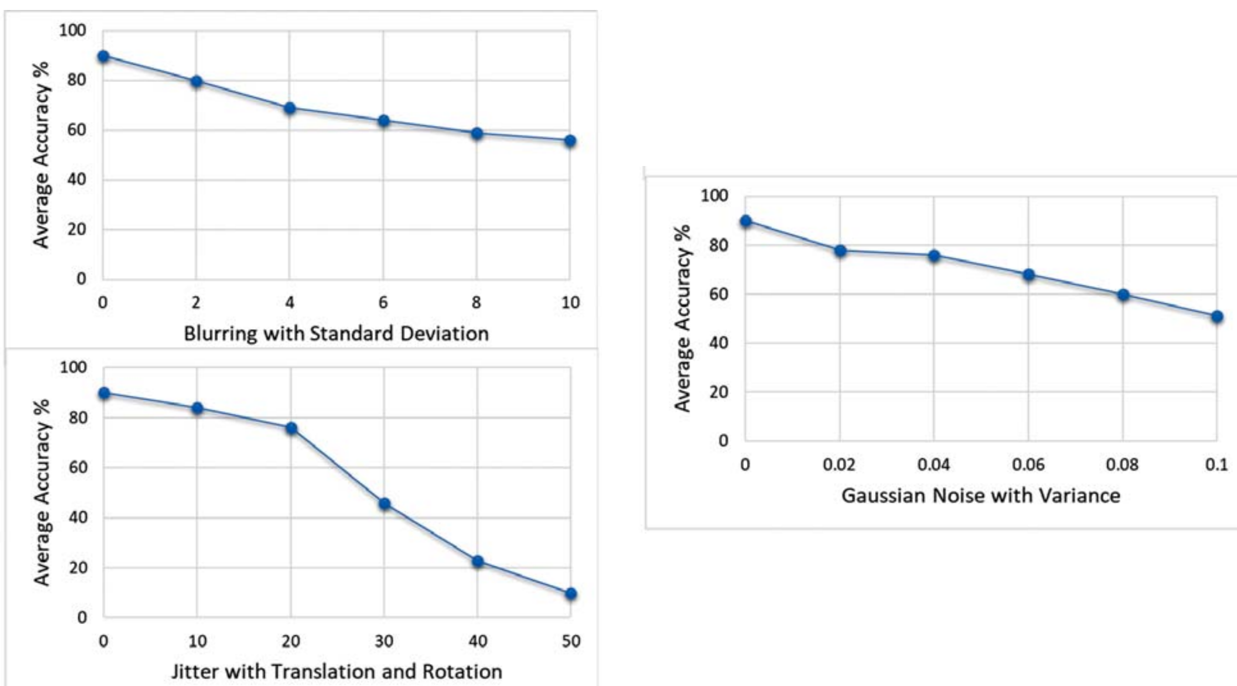


Fig. 7. Graphical Representation of Proposed System in Visual Noise.

Based on the classification results, only the viseme classes 5, 7, 8 and 9 have achieved an accuracy of more than 80%. This is due to their distinctive visual appearance. As the viseme classes corresponding to consonant phonemes are highly crowded (except for viseme 8 and 9), when compared to the viseme classes corresponding to vowel phonemes, the classifier fails to distinguish such similar visual appearances. In addition, the training dataset typically accounts for two-thirds of the overall dataset, with the remainder used for testing. As a result, this method may neglect some of the most informative data, resulting in a more significant bias. The solution to this problem is to use K-fold cross-validation. While splitting the dataset into K-folds, it is essential to ensure that each phoneme's relative class frequencies in each viseme are approximately preserved in each training and validation fold. Thus, the data in each class should be rearranged so that each fold is a valid representative of the whole. This process is termed stratification, and the associated cross-validation is stratified. Another main reason for the overall drop in performance is that the test set is utilized for both model selection and estimation. This overfits the test data, resulting in an optimistic bias in estimation. To overcome this issue, model selection and evaluation must be done separately, which is implemented using a modified version of cross-validation termed Nested Cross-Validation.

We adopt a modified version of SVM that addresses the issues of optimum hyperparameters, overfitting, and unbalanced datasets by using nested stratified 5-fold cross-validation (K. T. Kumar 2021) such that each observation serves as either training data, validation data, or test data. Nested cross-validation overcomes the issue of overfitting as in cross-validation, where the same test set is used for both model selection and estimation, resulting in a biased outcome. Stratified 5-fold cross-validation rearranges the data in each class, thereby ensuring that each fold is a reliable representative of the whole. The performance of the proposed VSR system in a clean visual environment based on a modified SVM classifier is shown in Figure 6.

The experimental results show that viseme recognition using a modified SVM classifier gives a better result, with an average of 90% in all metrics. Table 3 and Figure 6 display the performance of the proposed system in corrupted visual speech.

Jitter corruption drastically affects the viseme recognition rate when compared to the other types of noise. Due to the selection of relevant visual features and the proposed SVM classifier, the proposed system has reasonable noise robustness in blur and Gaussian noise. DCT has better noise robustness towards blurring and Gaussian noise even at higher noise levels, where the visual attributes are incomprehensible to human perception.

6. CONCLUSION

This paper investigates the effect of visual noise on recognizing the Malayalam viseme and seeks to make the system robust against noise using a modified SVM classifier. Proper linguistic knowledge is the backbone behind the success of any speech-based application. In this work, phoneme-to-viseme mapping and the identification of relevant frames from the video of the underlying phoneme are carried out by linguistic experts. Three real-time visual noises of varying intensities are used in this study. A nested stratified 5-fold cross-validation approach is used to select an optimum value for the hyperparameters of the SVM classifier and evaluate the model's performance, which minimizes overfitting issues and imbalances in the distribution of target classes. The proposed system produces an overall better outcome in accuracy, precision, recall, and F1-score for clean visual speech. The performance of the viseme recognizer degrades dramatically in jitter noise and less in blur and Gaussian noise. The results and the proposed implementation approach offer potential scope for improvement in the performance of speech systems against an extremely visual and acoustic noise background in a real-time scenario.

7. REFERENCES

[1] Aida-Zade, Kamil, Anar Xocayev and Samir Rustamov. 2017. "Speech Recognition Using Support Vector Machines." *Application of Information and Communication Technologies, AICT 2016 - Conference*

Proceedings **1**. https://doi.org/10.1109/ICAICT.2016.7991664.

[2] Bear, Helen L. and Richard Harvey. 2016. "DECODING VISEMES□: IMPROVING MACHINE LIP-READING Helen L . Bear and Richard Harvey." *Icassp* 2016, 2009-13.

[3] ---. 2017. "Phoneme-to-Viseme Mappings: The Good, the Bad, and the Ugly." *Speech Communication* **95,** 40-67. https://doi.org/10.1016/j.specom.2017.07.001.

[4] Bibish Kumar, K.T., Sunil John, K.M. Muraleedharan and R.K. Sunil Kumar. 2021. "Linguistically Involved Data-Driven Approach for Malayalam Phoneme-to-Viseme Mapping." *Applied Speech Processing,* pp. 117-45. https://doi.org/10.1016/b978-0-12-823898-1.00003-5.

[5] Fisher C.G. 1968. "Confusions among Visually Perceived Consonants." *Journal of Speech and Hearing Research,* **11**(4), 796-804.

[6] K. T. Bibish Kumar, R. K. Sunil Kumar, E. P. A. Sandesh, S. Sourabh and V. L. Lajish. 2019. "Viseme Set Identification from Malayalam Phonemes and Allophones." *International Journal of Speech Technology,* **22,** 1149-66.

[7] K. T. Kumar Bibish. 2021. "Viseme Identification and Analysis for Recognition of Malayalam Speech in Intense Background Noise." http://hdl.handle.net/10603/368521.

[8] Potamianos, Gerasimos, Etienne Marcheret, Youssef Mroueh, Vaibhava Goel, Alexandros Koumbaroulis, Argyrios Vartholomaios and Spyridon Thermos. 2017. "Audio and Visual Modality Combination in Speech Processing Applications." *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations,* **1,** 489-543. https://doi.org/10.1145/3015783.3015797.

[9] Skrelin Pavel A. 1999. "Allophone-Based Concatenative Speech Synthesis System for Russian." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* **1692,** 156-59. https://doi.org/10.1007/3-540-48239-3_28.

[10] Taylor Sarah L., Moshe Mahler, Barry John Theobald and Iain Matthews. 2012. "Dynamic Units of Visual Speech." *Computer Animation*, 2012 - ACM SIGGRAPH / Eurographics Symposium Proceedings, SCA 2012, 275-84.

The Journal of Acoustical Society of India : Vol. 50, No. 3-4, 2023 (pp. 138-147)

Perceptual and neural correlates of Multi-Verb constructions in Bengali language

Shankha Sanyal[1,2], Samir Karmakar[1], Shabaz Haque[1], Rajanya Purkait[1], Pijush Kanti Gayen[1], Sayan Nag[3] and Dipak Ghosh[2]

[1]School of Languages and Linguistics, Jadavpur University, Kolkata, India

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

[3]Department of Medical Biophysics, University of Toronto, Canada

e-mail: ssanyal.ling@jadavpuruniversity.in

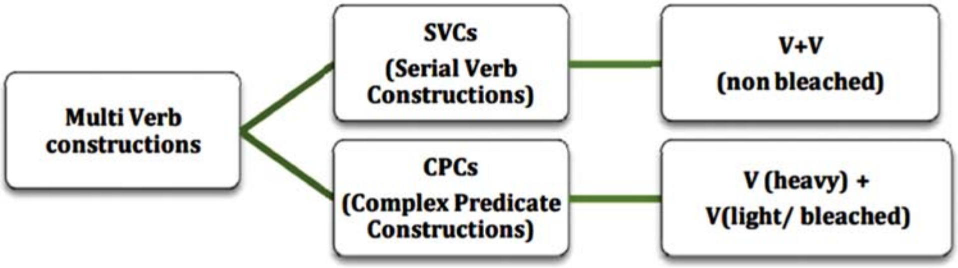
[Received: 27-03-2022; Revised: 01-06-2022; Accepted: 15-06-2022]

ABSTRACT

The world of our knowledge is full of events and their construals. Provided the nature of reality, man is always tasked with the urgency of construing the multitude of events that passes his way every single cognizing moment. One of the major goals of this study is to understand the events that are construed by the Bengali language speakers (both native and non-native) with the use of multi-verbs in words as well as at the sentential level. Event construals include the syntactic-semantic properties of the certain languages. In this study, we envisage to study the neuro-psychological cues associated with the processing of two specific types of multi-verb constructions, namely Complex Predicate Constructions (CPC) and Serial Verbs Constructions (SVC) in Bengali language. For the psychological research work, we did an experiment on 100 people (73 L1 Bengali Speakers and 27 L2 Bengali Speakers. For the neural research work, an experiment with 100 sentences was prepared using the Google questionnaire and given as auditory stimuli to 10 participants. The stimulus in the experiment included jumbled Serial Verbs (SV) and Complex Predicates (CP) in the form of words and sentences. We attempted to assess robust neurological features such as Relative Power (alpha- and theta-power). The single trial EEG demonstrated how the brain processes CPCs and SVCs and served as the foundation for research into event construals of multiverbs in Bengali. The EEG data was analysed by comparing alpha and theta power, which corresponded to CPCs and SVCs. The present study provides new and interesting information regarding mental processing of multi-verb constructions in Bengali language.

1. INTRODUCTION

Multi-verb constructions occur very frequently in most South-Asian languages. A multi-verb construction usually refers to a V+V predicational element, where the first verb helps in providing semantic information to the sentence and it is followed by a second verb which provides either aspectual (lexical) meaning or the original semantic interpretation of the verb itself. In this work, we focus our attention to multi-verb formations in Bengali language, which are essentially categorized into two main classes - Complex Predicates (CPs) and Serial Verbs (SVs)[1-4].



ʃe	boi-ʈa	rekʰe	dʝ-l-o	(CPC)
3.SG	book-CLF	keep-CP	give-PST-3	

"He kept the book."

In this CPC, the two verbal elements, forming the compound, refers to a single event in unison. The second verb in the complex, *dʝ-l-o* is a light verb with much of its normal semantic content bleached, and gives no lexical meaning apart from contributing aspectual information to the complex.

ʃe	boi-ʈa	rekʰe	e-l-o	(SVC)
3.SG	book-CLF	keep-CP	come-PST-3	

"He kept the book and came back."

Contrarily to that in case of CPC, in this SVC, there is no bleaching of semantic contents for either of the verbal elements in the compound. Hence, it refers to a series of events occurring sequentially. This denotes that *rekʰe* and *e-l-o* refers to two different events which retain their own individuality in the course of the action.

The Serial Verbs present in words or sentences might indicate two events in a sequence, whereas there are some events which can be interpreted both as SV and CP events due to the existing ambiguity. The psychological analysis revealed how our mind processes the CPCs and SVCs and provided the basis of study for the event construals of multi-verb constructions for both L1 and L2 users of Bengali language. A confusion matrix was developed to have an idea about the accuracy percentage of identification of serial verbs and complex predicate constructions in Bengali language. The mis-match matrix helped us to draw a statistical comparison between the native and non-native speakers of the language, how they understand and perceive it. Unlike other studies on multi-verbs construction which generally deals with the syntax, semantics, paradigmatic etc.[5-9], the present study analyses the verbal constructions through a psychological point of view.

To begin with the experimental work, at first, we had made a list of serial verbs and complex predicates from our nearby resources like daily life conversation, books, magazines, stories, novels, various research papers etc. and then arranged the same in an excel sheet. After that we had made two google forms, one for word level analysis and the other for sentence level analysis of Bengali event construals. Then we set off to collect feedback from participants of different social, cultural, economical, ethnic and linguistic backgrounds. The next step was to further process the collected data and analyse the results gained through the Mis-match matrix. Which we are going to discuss further, in detail, in the results and findings section.

Whereas, the main focus of neurological linguistics is to understand language by unfolding its underlying neural mechanisms. EEG experiments are conducted to detect activity patterns during the processing of various tasks involving language processing (multi-verb constructions in this study). ERPs have been extensively used in neuroscience since decades and several characteristic components have been identified which specifies neural properties relating to tasks executing language processing[2]. EEG produces rhythmic signals with distinct spectral and phase properties. We know that the brain produces neural oscillations at various frequency bands in different psychological states[10, 11]. These oscillations are called brain waves and can be measured using band pass filters. The best-known frequency bands are alpha (7.5 - 12.5 Hz) which represents a relaxed state of mind, beta (13-30 Hz) represents planned task execution, theta (4-8 Hz) normally appear in childhood and / or different sleep stages[12-14]. The parietal event related coexistence occurred in the theta band whereas the left hemispheric cortical areas record the alpha-band desynchronization.

The goal of this study is to approach the above theories through both psychological and neural point of view. At first by analysing Serial verbs and Complex predicates in an event based linguistic framework through the analysis of confusion matrix, cognitive response of the participants for CPC and SVC in the word and sentence level, perceptual reaction of the L1 and L2 users of Bangla Language and then by observing the average perceptual reaction time, the accuracy rate of perceiving a CPC as a CPC and *viz.* and finally a comparative study between the alpha power and theta power corresponding to CPC and SVC. At the end we draw the outcome of our research work by placing the results of both the psychological and neural analysis side by side. The main objectives of the study are listed as under:

1. To investigate the psychological and neural correlates of multi-verb processing in Bengali.
2. To investigate how multi-verb processing is perceived at the word and sentence levels using human response and EEG data from L1 and L2 Bengali speakers.
3. To examine the variations in alpha and theta power corresponding to processing of Serial Verbs and Complex predicates in Bangla and correspondingly the excitation levels in the different lobes of brain while processing the two distinct categories of multi-verbs.

2. METHODOLOGY

Keeping these basic characteristics of Complex Predicates and Serial Verbs in mind we would like to proceed ahead with the experiment. For this research work, we began with the psychological analysis by doing an experiment on 100 people (73 L1 Bengali Speakers and 27 L2 Bengali Speakers). Using Google Slides and Google forms we had created quite a few feedback response forms for the participants of the study. All the participants were requested to share their responses through the google form response sheets. One corpora was created by selecting 50 serial verbs and 50 complex Predicates along with distractors (Fig. 1).

Fig. 1. A template from the list of complex predicates and serial verbs.

All of these word level inputs were randomly shuffled and then arranged in a Google form (Fig. 2). Along with the responses of the participants, the google form also collected information like the participants' age, gender, mother tongue, etc. which helped us to categorise the responses in comparison to their age, gender, mother tongue and the level of exposure to the target language.

Likewise, another corpora was also created in a similar pattern. The only difference is that it was based on a sentence level analysis, whereas the other google form was based on a word level analysis. As

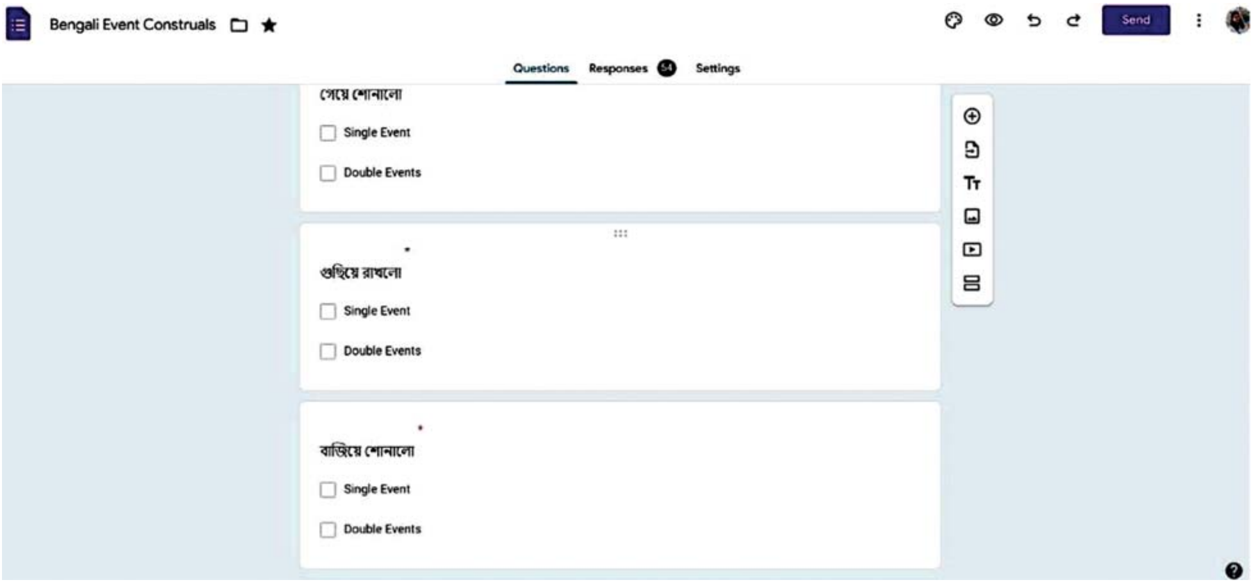


Fig. 2. A template from the Google form (Bengali Event Construals-1).

for the precious feedback of the Golden-Agers, we went to meet them personally and recorded their feedback. The best part of this experiment was the feedback collection, as we went door to door, to meet our friends, family and relatives. Along with the collection of the feedback, we had a great time talking to them and gaining wisdom from their experiences of life. After that, with great efforts and hard work of various days and nights, we sorted down all the collected data. Then updated the same in various excel sheets separately of the L1 and L2 Bengali speakers. Applying various formulas of calculation over the collected data then with continuous processing of the feedback received. The figures that we had gained were further arranged in a Mis-match Matrix. After analysing the output in the Mis-match matrix, the colour gradient gave us a clear idea of the results of the overall experiment. That we are about to discuss in detail in the next section.

After the psychological analysis, we performed the experimental work of neurological analysis. In this experiment, we prepared 150 sentences for the EEG dataset. The dataset for our analysis included a corpus of 150 sentences which have been categorized into Serial Verb Construction (SVC) and Complex

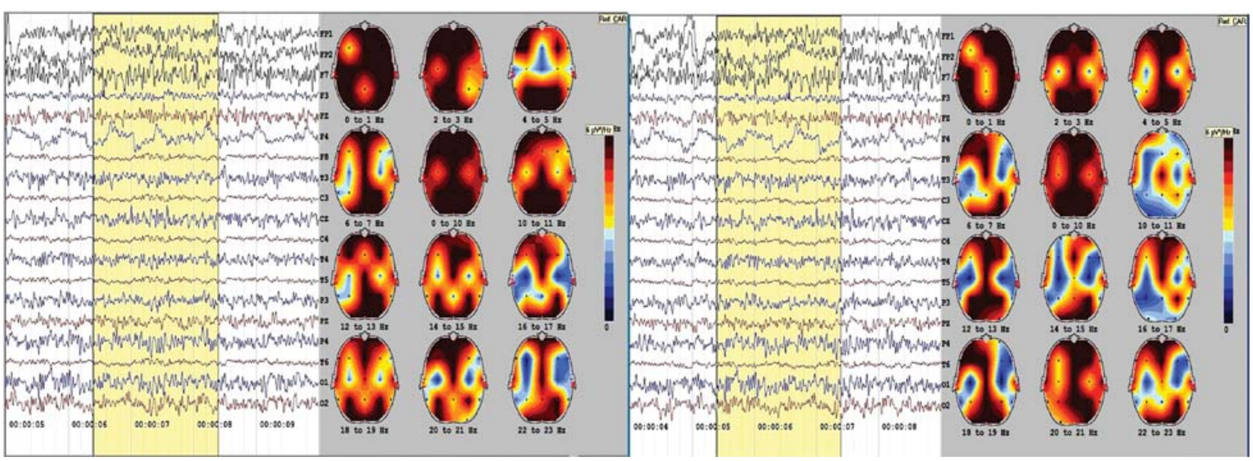


Fig. 3(a-b). Raw EEG data and frequency plots for 'rest state' and cognition of 'Serial Verbs'.

Predicate Constructions (CPC) based on event construals. The recordings were taken from 5 Male and 5 Female participants, who were asked to record the sentences provided along with a few distractors, which ensured the respondents are primed while recording the data. The participants were asked to respond to the auditory stimuli with EEG electrodes to monitor the neural activities. We examined EEG during specific brain states such as the resting state, serial verb and complex prediction conditions. Fig. 3 (a-c) denotes the real time EEG data and corresponding frequency plots for the three experimental sets we are analyzing.

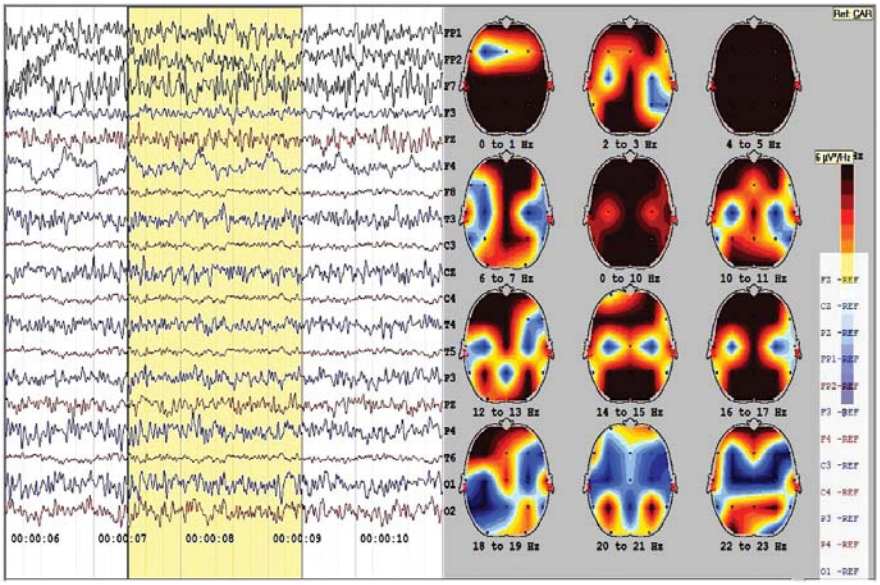


Fig. 3(c). Raw EEG data and corresponding frequency plots for 'Complex Predicates'.

This present study provides interesting and new information about the psycho-neural processing of complex predicates and serial actions. The analyses of the data are mentioned in the next section, where we will see the differences in the understanding of complex prediction and sequential action and relative power in the form of comparison between the alpha and theta powers of the corresponding CPC and SVC.

To the time series data {x_1, x_2,.......x_N}, we perform the Fast Fourier Transform (FFT), and the result obtained is denoted as {X_1, X_2, ...X_N}. A continuous frequency band from flow to fup is sliced into K bins which can be of equal width or not. Boundaries of bins are specified by a vector band = [f_1, f_2, ...f_K], such that the lower and upper frequencies of the i^{th} bin are fi and f_i+1 respectively. The bins used are δ (0.5 - 4 Hz), θ (4 - 7 Hz), α (8 - 12 Hz), β (12 - 30 Hz) and γ (30 - 50 Hz). For these bins we have band = [0.5, 4, 7, 12, 30, 50]. The power spectral intensity of the kth bin is evaluated as:

$$\mathrm{PSI}_k = \sum_{i=\left|N\left(\frac{f_k}{f_s}\right)\right|}^{\left|N\left(\frac{f_{k+1}}{f_s}\right)\right|} |X_i|^2 \tag{1}$$

k=1,2,....K-1

where f_s is the sampling rate, and N is the series length[15-17]. Our approach divides each 120 second data epoch into 8 windows, 30 sec wide with each window overlapping the previous window by 15 sec. Each window is converted into the frequency domain using Fast Fourier Transform (FFT). The frequency descriptors of the power bands, theta and alpha rhythms are extracted. The average power corresponding to each experimental condition was computed for all the electrodes of all lobes. The error bars give the SD values computed from the different values of spectral power for each electrode.

3. RESULTS AND DISCUSSION

Now, as we proceed towards the data analysis part, beginning with the psychological analysis, we know that the human response data taken from the 100 participants were analysed in the form of age, sex and other demographics. The participant pool consisted of L1 Bengali and L2 Bengali speakers. Table 1 and 2 shows the confusion matrix corresponding to the 'word' class and 'sentence class' respectively.

Table 1. Confusion Matrix corresponding to 'word class' of Serial Verbs and Complex Predicates.

			Predicted	
			Single	**Double**
Bengali	**Target**	**Single**	79.92156863	24.2745098
		Double	46.15384615	53.80769231
			Predicted	
			Single	**Double**
Non-Bengali	**Target**	**Single**	70.03921569	26.96078431

Table 2. Confusion Matrix corresponding to 'sentence class' of Serial Verbs and Complex Predicates.

			Predicted	
			Single Event	Double Event
Bengali	Target	Single Event	63.33	
		Double Event	38.57104762	61.43
			Predicted	
			Single Event	Double Event
Non-Bengali	Target	Single Event	49.99	50
		Double Event		67.14

In the word level analysis, Both L1 and L2 speakers identify 'single events' or complex predicates more accurately as compared to 'double events'(serial verbs). In case of Non-Bengali speakers, we see that the Accuracy percentage is much lower corresponding to the identification of complex predicates. In case of serial verbs, however, an interesting finding is that the L2 Bengali speakers have higher accuracy percentage as compared to L1 Bengali speakers. This shows that the perception of serial verbs is higher in case of Non-Bengali participants. Although the perception of serial verbs for both the cases has much lower Accuracy percentage.

In the sentence level analysis, for L1 participants the perception rate is more than 60% in case of both single event and double event. Whereas, the percentage of misconception for L1 participants is higher in the case of double events as compared to single events. And the ambiguity for L1 speakers gets resolved, when shifted from word class to sentence class, whereas, L2 participants could identify Serial verbs way better than Complex predicates.

Therefore, for the sentence class and word class corpus, we observed distinctly different perceptual abilities in the above section. Thus, we can conclude by saying that, if we consider the Bengali knowing population, in that case the ambiguity (considering the miss classification as ambiguity) gets resolved to a greater extent when they are shifted from word class to sentence class.

From Fig. 4, we have an idea about the functionalities of different lobes of the brain and we focus on frontal, parietal and temporal lobes for this particular work. Fig. 5 shows the EEG alpha and theta power variation for the frontal lobe as per Eqn. (i) in the methodology section.

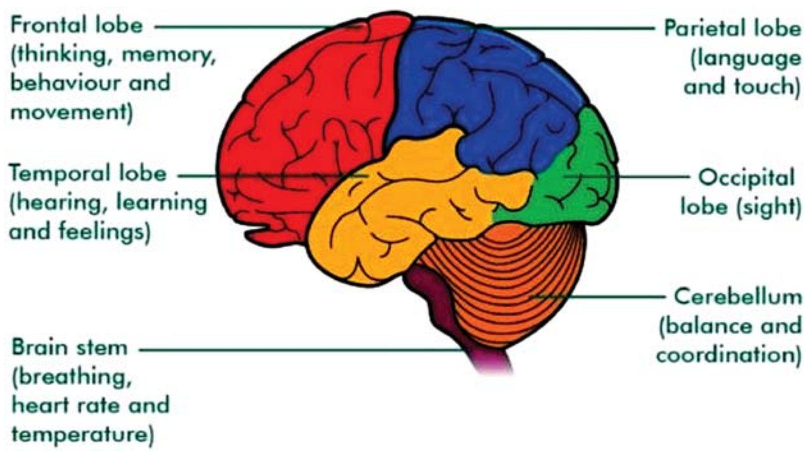


Fig. 4. Functions of different lobes of the human brain.

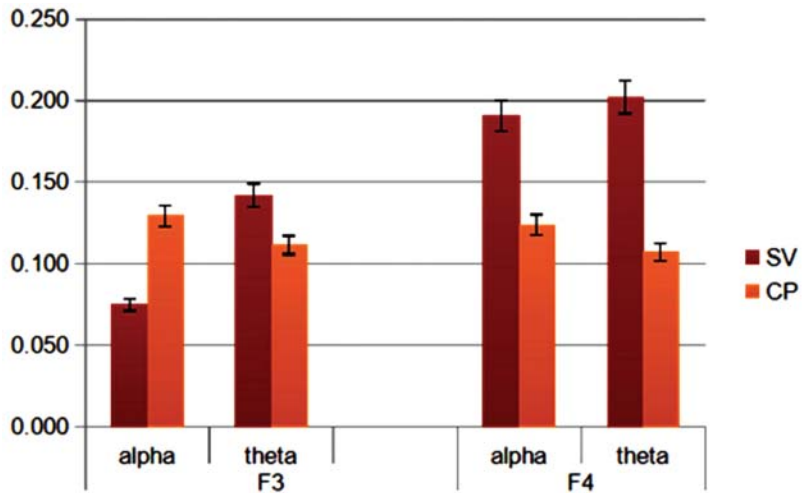


Fig. 5. Variation of alpha and theta power corresponding to SV and CP in the frontal lobe.

As we shift our focus towards the neural analysis part, for the frontal lobe, we see an effect of right-left lateralization in the processing of serial verbs and complex predicates corresponding to alpha power - this can be ascribed to the alpha asymmetry and cognitive load distinction in the frontal lobe. While for serial verbs, we have lower alpha power in the left hemisphere, higher alpha power is seen in the right frontal lobe. This can be attributed to higher cognitive load associated with the processing of serial verbs. In the theta frequency domain, for both the left and right frontal lobe, higher theta power is seen corresponding to the processing of Serial verbs. Thus higher cognitive load corresponds to higher theta power in both the hemispheres of the front lobe. Fig. 6 shows the variation for the parietal lobe.

For the parietal lobe, we see for the right and left electrodes, lower alpha power corresponds to the processing of Serial verbs, while for complex predicates alpha power is higher. The situation is exactly the reverse for theta power. Theta power is considerably higher while processing serial verbs as compared to processing complex predicates in both right and left parietal electrodes. Here a specific distinction in neural processing is observed for the pair of electrodes. Fig. 7 shows the variation of alpha and theta power corresponding to the temporal lobe.

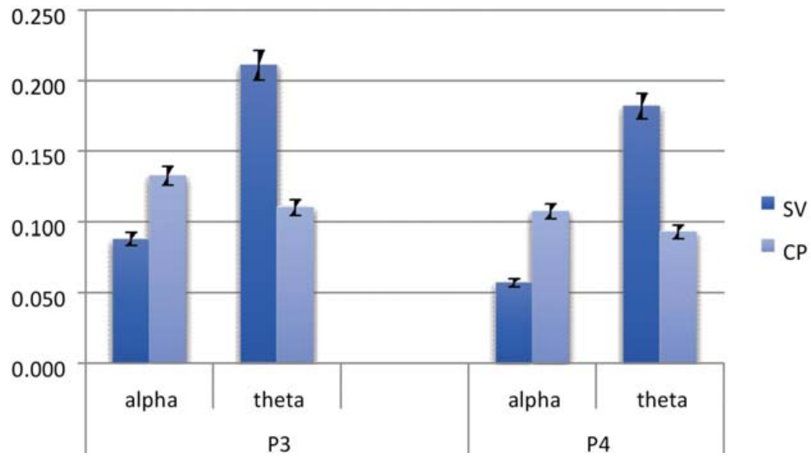


Fig. 6. Variation of alpha and theta power corresponding to SV and CP in the parietal lobe.

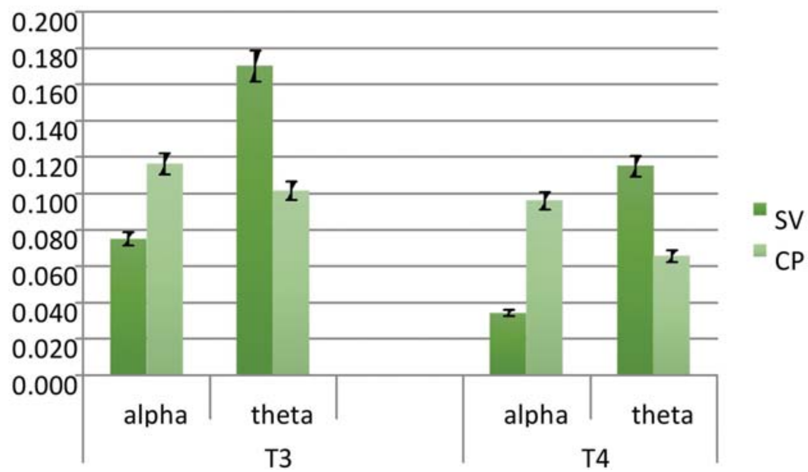


Fig. 7. Variation of alpha and theta power corresponding to SV and CP in the temporal lobe.

In the study of Temporal lobe, an effect of both hemispheres has been seen in the path of studying serial verbs and complex predicates corresponding to alpha power - this can be ascribed to the alpha asymmetry and cognitive load distinction in the temporal lobe. Alpha power is low in both the left and right temporal lobes compared to theta power in serial verb processing. Theta power is high in both left and right temporal lobes in the course of serial verb processing.

4. CONCLUSION

The present study makes use of confusion matrices, as part of human response analysis for both groups of speakers, while the neural part makes use of power spectrum EEG responses generated from different lobes of the brain. The following interesting conclusions can be made from the study.

1. Both Bengali and Non-Bengali Subjects identify single events with a higher accuracy percentage as compared to double events in the 'word class', in the 'sentence class', however the ambiguity is resolved to a large extent in case of 'double events', as the accuracy percentage for serial verbs increases considerably in the 'sentence class' as compared to word class.
2. Another interesting finding is that L2 Bengali speakers correctly identify 'serial verbs' belonging to 'word class' and 'sentence class' to a larger extent as compared to Bengali speakers.

3. Alpha asymmetry is observed for the processing of cognitive loads associated with Serial verbs and Complex predicates construalization in the frontal lobe. A clear case of neural classification is seen for the processing of two different types of event construals in the frontal lobe for Bengali language.
4. In the parietal and temporal lobe, we see distinct patterns of alpha and theta EEG spectral power variation corresponding to Serial Verbs and Complex predicate process. While alpha power remains low for Serial verb processing, theta power remains high for both the electrodes.

The future prospects of this work are as follows,

- o The present pilot study can be extended to a larger participant pool with greater variety of word and sentence variations to yield a more conclusive result.
- o Considering more areas of the brain a definite response mapping corresponding to speech level analysis of Bengali multi-verbs is being done as part of this work.

5. REFERENCES

[1] Puttaswamy, C. 2018. Complex predicates in South Asian languages: an introduction. *JSALL*; **5**(1), 1-3, 1

[2] Basu D. and Wilbur R. 2010. Complex predicates in Bangla: An event-based analysis. *Rice Working Papers in Linguistics,* **2,** 1-19.

[3] Das D., Pal S., Mondal T., Chakraborty T. and Bandyopadhyay S. 2010. Automatic extraction of complex predicates in Bengali. *In Proceedings of the 2010 Workshop on Multiword Expressions: from Theory to Applications,* pp. 37-45.

[4] Bhattacharja S. 2018. On the formation of complex predicates in Bengali. *Journal of South Asian Languages and Linguistics,* **5**(1), 5-21.

[5] Karmakar S. and Kasturirangan R. 2010. Cognitive processes underlying the meaning of complex predicates and serial verbs from the perspective of individuating and ordering situations in bānlā. *In Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia,* pp. 81-87.

[6] Yadav B.K. 2022. *Complex predicates in Maithili* (Doctoral dissertation, Faculty of Linguistics).

[7] Abbi A. and Gopalakrishnan D. 1991. Semantics of explicator compound verbs in South Asian languages. *Language Sciences,* **13**(2), 161-180.

[8] Butt M. 1997. *Complex predicates in Urdu.*

[9] Ramchand G. 1990. *Complex predicate formation in Bengali.*

[10] Chowdhury S., Gayen P.K., Sarkar U., Banerjee A., Sanyal S., Karmakar S. and Ghosh D. 2021. An acoustic exploration of event construals in Bengali language. *In Journal of Physics: Conference Series, IOP Publishing.,* **1896**(1), 012012.

[11] Buzsaki G, D. A. 2004. *Neuronal Oscillations in Cortical Networks, Science,* **304**.

[12] Gayen, P. K., Sanyal S., Banerjee A., Karmakar S. and Ghosh D. 2023. Oscillatory Neural Responses for Order and Request Speech Acts: An EEG Study on Bengali Utterances. *Available at SSRN 4351641.*

[13] Basu M., Sanyal S., Banerjee A., Nag S., Banerjee K. and Ghosh D. 2023. Neural quantification of timbre and emotions from Indian Classical Music: A multifractal exploration. *Physica A: Statistical Mechanics and its Applications,* **624,** 128937.

[14] Nag S., Basu M., Sanyal S., Banerjee A. and Ghosh D. 2022. On the application of deep learning and multifractal techniques to classify emotions and instruments using Indian Classical Music. *Physica A: Statistical Mechanics and its Applications,* **597,** 127261.

[15] Sarkar U., Pal S., Nag S., Sanyal S., Banerjee A., Sengupta R. and Ghosh D. 2021. A simultaneous EEG and EMG study to quantify emotions from hindustani classical music. *In Recent Developments in Acoustics: Select Proceedings of the 46th National Symposium on Acoustics, Springer Singapore,* pp. 285-299.

[16] Sanyal S., Banerjee A., Pratihar R., Maity A.K., Dey S., Agrawal V. and Ghosh D. 2015. Detrended fluctuation and power spectral analysis of alpha and delta EEG brain rhythms to study music elicited emotion. *In 2015 International Conference on Signal Processing, Computing and Control (ISPCC), IEEE,* pp. 205-210.

[17] Bishwamitra R. and Maringanti H.B. 2021. Frequency Analysis of Human Brain Response to Sudarshan Kriya Meditation. *In Data Science in Engineering and Management, CRC Press,* pp. 79-86.

The Journal of Acoustical Society of India : Vol. 50, No. 3-4, 2023 (pp. 148-155)

Relation between performance on psychoacoustical tests and rating of musical abilities in music trainees

Sindhu Parthasarathy, Asha Yathiraj and Sreelatha K.R.
JSS Institute of Speech and Hearing, Mysuru, Karanataka-570 004, India
e-mail: sindhup@jssish.in

[Received: 31-03-2022; Revised: 17-06-2022; Accepted: 01-07-2022]

ABSTRACT

Background: Music training is known to enhance auditory skills. However, studies reported in literature rarely focus on the duration of music training that can bring about an improvement in auditory skills. Such information would be useful to decide the minimum duration of providing music training to bring about any improvement in auditory perception. This would enable planning a music training program for those with an auditory processing disorder, a condition where the central nervous system is unable to utilise auditory information effectively and efficiently. While several training procedures have been used for those with auditory processing disorders, there is limited information regarding the use of music training as a form of rehabilitation for those with the condition. Hence, the study aimed to determine the duration of music training that can bring about improvement in psychoacoustical tests. The study also aimed to determine if a relation exists between the performance of music trainees on different psychoacoustical tests and the rating by their music teacher regarding their musical abilities.

Method : Fourteen children in the age range of 7 to 10 years, who had undergone Carnatic music training for less than 1 year, were studied. They were grouped as those with 2 to 6 months of music training (Group I) and those with 8 to 12 months of music training (Group II). Their 'difference limen for frequency,' 'gap detection threshold,' and 'duration discrimination' were evaluated. Their music teacher also rated their musical abilities on a five-point rating scale, which assessed pitch, rhythm, tempo, auditory memory, and overall musical abilities.

Results : The two participant groups did not differ significantly on any of the psychoacoustical parameters that were evaluated. Also, the participants obtained significantly better thresholds on most parameters that were tested compared to the available norms of those who had not undergone music training. However, no significant correlation was found between the responses on the psychoacoustical tests and the musical skill rating by the teacher.

Discussion : The study indicated that the children who underwent musical training performed significantly better in psychoacoustical tests when compared with those who did not undergo any music training, demonstrating that music training had a positive impact. The duration of music training did not affect their psychoacoustic perception, indicating that the improvement seen subsequent to music training can be observed within 6 months. Thus, if music training is to be provided to those diagnosed to have an auditory processing problem, improvement can be anticipated within 6 months. Further, the lack of correlation between the psychoacoustical test values and the rating by the music teacher indicates that the two probably measured different aspects of auditory perception.

1. INTRODUCTION

Music training is known to have a positive impact on cognitive functioning (Faßhauer, Frese and Evers 2015), vocabulary (Linnavalli *et al.* 2018), language skills (Mizener 2008), phonological awareness and reading (Anvari *et al.* 2002), social and communication development (Gerry, Unrau, and Trainor 2012), as well as performance on psychoacoustical tests (Jain *et al.* 2019). In children, extensive music training has been found to improve auditory discrimination of fine modulations of intensity, frequency and time (Kraus and Chandrasekaran 2010). Additionally, it has been observed to result in enhancement in temporal processing abilities (Ribeiro, Scharlach, and Pinheiro 2015), pitch discrimination abilities (Micheyl *et al.* 2006), auditory stream segregation (Beauvois and Meddis 1997), auditory attention (Strait *et al.* 2009), prosody processing (Wong *et al.* 2007), and long-term spatial-temporal reasoning (Rauscher *et al.* 1997).

The positive impact of music training on auditory abilities has been usually reported after at least 1 to 1.5 years of enrolment (Fujioka *et al.* 2006; Koelsch *et al.* 2005; Hyde *et al.* 2009). While a significant improvement on a digit span test was seen after one year of musical training (Fujioka *et al.* 2006), enhancement in auditory skills such as music listening and discrimination skills was noted as early as 15 months (Hyde *et al.* 2009). Jain, Mohamed, and Kumar (2014) found that eight sessions of training to identify ragas were not adequate to improve auditory skills (frequency discrimination, intensity discrimination, gap detection, temporal modulation transfer function and duration pattern). This implies that a longer duration of training is required to obtain changes in psychoacoustic tasks. However, the positive impact of short-term music training has been found through the use of objective measures such as electrophysiological auditory evoked potentials and electroencephalography. Lappe *et al.* (2011), using an event-related auditory evoked potential as well as magnetoencephalographic response, observed that practising rhythmic-focused piano exercises enhanced responses to signals that varied in rhythm and enhanced cortical activity. This was observed after 30 minutes of sensorimotor-auditory music training for 8 days over 2 weeks. Likewise, using electroencephalography, Carpentier, Moreno, and McIntosh (2016) reported that short-term musical training of just 20 days increased electroencephalography complexity for coarse temporal scales that were associated with global information processing.

While musical abilities have been found to improve psychoacoustical performance as well as brain activity, there is a dearth of literature on whether such enhancement can be measured using a perceptual judgement scale. Such a measure would make it possible to evaluate the auditory abilities of individuals undergoing music training rapidly, at regular time intervals. The individuals would not have to undergo frequent psychoacoustic tests, which are usually time-consuming, and the results of which may be compromised by the familiarity of the test material when utilised repeatedly. With the use of a rating scale by a music teacher, the trainees can be evaluated frequently to record their improvement. Further, Saunders and Holahan (1997) found rating scales for music to be useful in analysing the strengths and weaknesses of the performance of music trainees. It needs to be seen if there exists a relation between the perceptual rating by a teacher and the performance of the trainees on psychoacoustic tests such as duration discrimination, frequency discrimination, and temporal resolution. Additionally, these psychoacoustical tests are often used to measure improvement following music training. Additionally, such psychoacoustic tests have also been used to measure auditory processing abilities. The presence of a correlation between perceptual rating and psychoacoustic performance would enable the former to serve as an easy tool to measure improvement in auditory processing subsequent to music training. This would be useful not only in typically developing children but also in those diagnosed to have an auditory processing problem. Thus, the study aimed to establish whether a limited period of training can result in an improvement in psychoacoustical performance compared to existing normative values available for non-musicians. In addition, the study also aimed to determine the relation between the performance of music trainees on psychoacoustical tests and the teacher's rating of their musical skills.

2. METHODS

Subject selection : The study included 14 children (6 males & 8 females) aged 7 to 10 years (mean = 8.83 years, SD = 1.00), who had undergone Carnatic music training, a form of music taught in south India. The sample size was calculated using G*power (version 3.1.9.6), with a 0.05 error of probability and power of 0.75, based on the study by Sangamanatha (2012). The children underwent online or offline training for 45-minute twice a week, for periods ranging between 2 to12 months. The participants were divided into two groups based on the duration of training they had undergone. The participants in Group I (n = 6) underwent training for 2 to 6 months (mean = 4.36 months) and those in Group II (n = 8) underwent training for 8 to 12 months (mean = 10.88 months). Of the ten lessons that are usually taught in Carnatic music (sarale varse, thaara sthaayi, mandara sthaayi, janti varase, dhatu varase, alankara, geethegalu, jathi swara, varna, & keerthanegalu), Group 1 was taught up to lessons 3 or 5, and Group II was taught up to lessons 6 or 7. The children were taught to sing along with the fundamental note (shruthi), following the required beats (thaala). The children had no known hearing, speech-language problems, or any other associated problems.

Procedure : All the participants were evaluated using three psychoacoustical tests [Difference Limen for Frequency (DLF), Gap Detection Threshold (GDT), & Duration Discrimination Test (DDT)]. The psychoacoustical tests were performed in a quiet room, free from visual and auditory distractions. The noise levels in the room ranged between 37 to 40 dBA. Half the children were tested in their right ear and the other half in their left ear by an audiologist, to avoid an ear effect. The stimuli were played using MATLAB (version 7.10) maximum likelihood procedure toolbox, loaded on a personal computer. The volume control of the computer was manipulated such that the output through a Sennheiser HD206 headphone was 65 dB SPL. A three-interval-forced-choice method was used to obtain the thresholds for all the tests, where two of the stimuli were similar (standard stimulus) and one was different (variable stimulus). The participant had to identify which of the three stimuli was deviant. Depending on the response of the participants, the software automatically increased or decreased the difference between the standard and the variable signals. Thirty sets of stimuli were presented to track the threshold using an 80% correct response criterion.

DLF was obtained for three standard pure tones (500 Hz, 1000 Hz, & 4000 Hz). For each of the standard tones, blocks of three stimuli were presented, of which two had the standard frequency and the other had a frequency that was higher than the standard stimulus. The maximum deviation from the standard frequency tone was 10%. Each participant had to verbally respond by saying "first," "second," or "third" to indicate which of the three stimuli in a block contained the higher frequency. The responses were manually keyed-in by the examiner into the MATLAB toolbox.

GDT was measured by determining the smallest detectable gap in a 750 ms white noise. The standard stimulus was a 750 ms white noise with no gap and the variable stimulus had a gap that varied in duration from 0.1 ms to 64 ms. The participants had to identify which of the three stimuli in a block had the gap.

DDT was performed to establish the smallest difference in duration that could be discriminated in a triad of 1000 Hz tones that varied in duration. The standard stimulus had a duration of 250 ms, and the deviant tone was longer than the standard tone by 0.1 ms to 200.1 ms. The participant had to identify which of the three stimuli in a block was longer.

Teacher Rating : A 'Musical rating scale,' developed as a part of the study, was used by the music teacher to grade the perceptual abilities of the trainee on five domains (pitch, rhythm, tempo, auditory memory, & overall musical skills). The domains selected for the scale were based on the recommendations of two musicians and two audiologists with at least 8 years of experience. Each domain was assessed using a five-point rating scale, where one was the lowest score and five was the highest score (Fig. 1). The maximum possible rating score was 25. The music teacher assessed each child twice, with a gap of five days between the two ratings, to check test-retest reliability.

Analyses : The data were subjected to statistical analyses using SPSS (Version 21). Both descriptive and inferential statistics were carried out. As the data were not normally distributed on a Shapiro-Wilks

Fig. 1. Musical Rating Scale.

test ($p < 0.05$), nonparametric statistics were used. In addition, parametric statistics, which is known to be more powerful, was also done.

3. RESULTS AND DISCUSSION

Initially, the test-retest reliability of the rating of the music teacher was checked using a Cronbach Alpha (α). It was found to be 0.90, indicating good test-retest reliability. Also, there was no significant difference between the total rating score obtained across the two evaluations, measured using a Wilcoxon signed rank test ($z = -1.342$, $p = 0.18$). A paired t-test also revealed that there was no significant difference between the total scores rating of the music teacher across the two evaluations [$t(13) = -1.385$, $p = .189$]. Hence, further analyses were done with the values of the first rating.

The median, inter-quartile range, mean, and standard deviation of the psychoacoustical tests performed are provided separately for each of the groups in Table 1. From the table, it is evident that the median thresholds of those with > 6 months of training and those with ≤ 6 months of training varied depending on the stimulus as well as the test that was administered. A Mann Whitney U test indicated that there

Table 1. Median and mean thresholds of the psychoacoustical tests with inter-quartile range / standard deviation given in parenthesis.

Tests	Standard Stimuli	Median values (Inter-quartile Range)			Mean (Standard deviation)		
		Groups I + II (n = 14)	Group I (n = 6)	Group II (n = 8)	Groups I + II (n = 14)	Group I (n = 6)	Group II (n = 8)
DLF	500 Hz	35.92 (18.54)	32.55 (19.44)	35.92 (20.26)	34.95 (10.95)	34.76 (10.92)	35.10 11.72
	1000 Hz	47.14 (68.39)	66.84 (70.39)	47.14 (66.82)	62.16 (33.15)	65.09 (37.60)	59.97 (31.90)
	4000 Hz	177.95 (239.57)	209.49 (264.42)	177.95 (155.66)	193.95 (112.44)	212.10 (140.23)	180.34 (94.56)
GDT	750 ms white noise with no gaps	2.73 (0.82)	2.73 (3.45)	2.73 (0.99)	3.42 (3.04)	4.47 (4.62)	2.63 (0.52)
DDT	1000 Hz tone with 250 ms duration	82.41 (37.4)	78.51 (29.01)	94.3 (38.8)	88.97 (22.40)	83.26 (21.35)	93.25 (23.60)

Note: Group I = Participants with ≤ 6 months of training; Group II = Participants with > 6 months of training was no significant difference between those who had training for ≤ 6 months and those who had training for > 6 months for DLF for 500 Hz (U = 23.5, p = .95), 1 kHz (U = 24, p = 1), and 4 kHz (U = 20, p = .66); as well as DDT (U = 17, p = .41), and GDT (U = 21, p = .75). An independent sample t-test also confirmed that there existed no significant difference between the two groups for DLF for 500 Hz [t(12) = -.05, p = .96], 1 kHz [t(12) =.276, p =.79], and 4 kHz [t(12) = .51, p = .62]; DDT [t(12) = -.82, p = .43]; and GDT [t(12) = 1.13, p = .28].

Further, a Spearman's rank correlation was performed to check the relation between the psychoacoustical test performance of the participants and the rating given by the music teacher. No significant correlation was found between the results of the three psychoacoustical tests and the rating given by the teacher (p > 0.05) for each of the domains of the music rating scale. Similar results were obtained when a Pearson correlation was administered (p > 0.05).

To confirm whether the music training brought about an improvement in psychoacoustical performance, the values obtained by the participant groups were compared with existing norms of individuals with no music training using a one-sample t-test. The comparison was done with data from the two groups combined as they did not differ significantly. The DLF thresholds, measured at 500 Hz, 1 kHz, and 4 kHz, were compared with the mean values obtained by Mithlaj (2018). Likewise, the values obtained for GDT and DDT were compared with that obtained by Prithvi (2018) and Barman (2007), respectively. As can be seen in Table 2, there was a significant difference for most parameters studied between the participants of the current study who had undergone music training and the norms available in the literature for those who had not undergone music training.

Table 2. Comparison of thresholds obtained by the participants who had undergone music training with norms available in the literature for those who had not undergone music training for Difference Limen for Frequency (DLF), Gap Detection Threshold (GDT), and Duration Discrimination Test (DDT).

Tests	Standard Stimuli	Median values (Inter-quartile Range)		
		t	*df*	*p*
DLF	500 Hz	–22.23	13	0.001
	1000 Hz	–4.27	13	0.001
	4000 Hz	1.46	13	0.167
GDT	750 ms white noise with no gaps	–2.75	13	0.016
DDT	1000 Hz tone with 250 ms duration	–7.41	13	0.001

The significantly better thresholds obtained by the children of the current study who had undergone music training compared to those individuals who did not have music training highlights the positive impact of music on psychoacoustic performance. This significantly higher performance in the children who had undergone music training was evident despite their thresholds for all three tests (DLF, GDT, & DDT) being compared with values obtained by those who were older than them. It is well documented that with maturation improvement in thresholds occurs for psychoacoustical tests (Elfenbein, Small, and Davis 1993; Jain 2016; Moore *et al.* 2008; Sangamanatha *et al.* 2012; Sutcliffe and Bishop 2005). The comparison was done with those who were older as normative values were not available for those in the age range evaluated in the current study. Similar to what was seen with the mean scores of the participants who underwent music training was also seen in their individual scores. This shows that music training led to improved psychoacoustical performance, not just when the group data were compared, but also when the individual scores were compared.

The findings of the study also indicated that the performance of the children on the psychoacoustical tests was similar in those who had ≤ 6 months of music training and those who had > 6 months of training. This implies that improvement in psychoacoustical test performance that occurs subsequent to musical training is evident within six months of training. Past studies on the positive effects of short-term music training have proven that limited training can bring about improvement when measured using objective measures such as mismatch negativity, magnetoencephalographic response, and electroencephalography (Lappe *et al.* 2011; Carpentier, Moreno, and McIntosh 2016). However, in the literature, it is reported that the positive effect of short-term music training (eight sessions) on psychoacoustic tests is not evident when behavioural measures are used (Jain, Mohamed, and Kumar 2014). Unlike what was reported by Jain, Mohamed, and Kumar (2014), improvement in psychoacoustical test performance was evident in the participants of the present study. A few of the children showed this improvement subsequent to music instruction after just eight sessions. The music teacher also reported that depending on the abilities of each child, the improvement was observed after 2 to 8 sessions. From this, it can be inferred that children with an auditory processing disorder may benefit from short-term music training, well within 6 months.

In addition, the findings of the current study indicated that the judgement of the perception of musical skills by the music teacher of the children was not related to their performance on the different psychoacoustical tasks. It is possible that the teacher's rating of the different domains was influenced by perceptual parameters such as timber, loudness, roughness, harmony, beat strength, rhythmic regularity, as well as the effort or energy of the voice. Such parameters have been noted to influence the perception of music (Friberg and Hedblad 2011; Friberg, Schoonderwaldt, and Hedblad 2011). Thus, it is possible that the teacher's rating and the psychoacoustic tests evaluated different aspects of music. However, the lack of association between the teacher's rating and the psychoacoustical test performance could have occurred as only a 5-point rating scale was used, resulting in the teacher rating most of the children in a similar manner. To avoid this, it is recommended that instead of a 5-point rating scale, a wider rating scale be used to help the music teacher differentiate the musical abilities of the children. Such differentiation may result in a better association between the rating of the teacher and the psychoacoustical tests.

Thus, the findings of the study indicate that those who underwent musical training for a shorter duration performed similarly to those who were trained for a longer duration, indicating that improvement in psychoacoustical performance is evident within a short period of music training. Based on these findings it can be surmised that providing training using a recreational activity such as music to those diagnosed to have an auditory processing problem may help improve their auditory perception within a limited time.

5. REFERENCES

[1] Anvari Sima H., Laurel J. Trainor, Jennifer Woodside and Betty Ann Levy. 2002. "Relations among musical skills, phonological processing, and early reading ability in preschool children." *Journal of experimental child psychology,* **83**(2), 111-130.

[2] Barman Animesh. 2007. "Psycho-Acoustical profile in normals and individuals with auditory dys-synchrony." Thesis: *All India Institute of Speech and Hearing.*

[3] Beauvois, Michael W, and Ray Meddis. 1997. "Time decay of auditory stream biasing." *Perception psychophysics,* **59**(1): 81-86.

[4] Carpentier Sarah M., Sylvain Moreno and Anthony R. McIntosh. 2016. "Short-term music training enhances complex, distributed neural communication during music and linguistic tasks." *Journal of Cognitive Neuroscience,* **28**(10), 1603-1612.

[5] Elfenbein Jill L., Arnold M. Small and Julia M. Davis. 1993. "Developmental patterns of duration discrimination." *Journal of Speech, Language, Hearing Research,* **36**(4), 842-849.

[6] Faßhauer Caroline, Achim Frese and Stefan Evers. 2015. "Musical ability is associated with enhanced auditory and visual cognitive processing." *BMC Neuroscience,* **16**(1), 1-9.

[7] Friberg Anders and Anton Hedblad. 2011. "A comparison of perceptual ratings and computed audio features." *Proceedings of the 8th sound and music computing conference.*

[8] Friberg Anders, Erwin Schoonderwaldt and Anton Hedblad. 2011. "Perceptual ratings of musical parameters." Measured Interpretation-Computer-based Performance Analysis in an Interdisciplinary Cross-examination. *Schott: Mainz,* pp. 237-253.

[9] Fujioka Takako, Bernhard Ross, Ryusuke Kakigi, Christo Pantev and Laurel J. Trainor. 2006. "One year of musical training affects development of auditory cortical-evoked fields in young children." *Brain*, **129**(10), 2593-2608.

[10] Gerry David, Andrea Unrau and Laurel Trainor. 2012. "Active music classes in infancy enhance musical, communicative and social development." *Developmental Science,* **15**(3), 398-407.

[11] Hyde Krista L., Jason Lerch, Andrea Norton, Marie Forgeard, Ellen Winner, Alan C. Evans and Gottfried Schlaug. 2009. "Musical training shapes structural brain development." *Journal of Neuroscience,* **29**(10), 3019-3025.

[12] Jain Chandni. 2016. "Relationship among psychophysical abilities, speech perception in noise and working memory on individuals with normal hearing sensitivity across different age groups." Thesis: *All India Institute of Speech and Hearing, Mysuru.*

[13] Jain Chandni, N. Devi, Sindhu Parthasarathy and S. Kavitha. 2019. "Effect of musical training on psychophysical abilities and working memory in children." *Journal of Indian Speech Language and Hearing Association*, **33**(2), 71-74.

[14] Jain Chandni, Hijas Mohamed and Ajith U. Kumar. 2014. "Short-term musical training and pyschoacoustical abilities." *Audiology Research,* **4**(1), 102.

[15] Koelsch S., T. Fritz, K. Schulze, D. Alsop and G. Schlaug. 2005. "Adults and children processing music: an fMRI study." *Neuroimage,* **25**(4), 1068-76.

[16] Kraus Nina and Bharath Chandrasekaran. 2010. "Music training for the development of auditory skills." *Nature Reviews Neuroscience,* **11**(8), 599-605.

[17] Lappe Claudia, Laurel J. Trainor, Sibylle C. Herholz and Christo Pantev. 2011. "Cortical plasticity induced by short-term multimodal musical rhythm training." *PloS one,* **6**(6), e21493.

[18] Linnavalli Tanja, Vesa Putkinen, Jari Lipsanen, Minna Huotilainen and Mari Tervaniemi. 2018. "Music playschool enhances children's linguistic skills." *Scientific Reports,* **8**(1), 1-10.

[19] Micheyl Christophe, Karine Delhommeau, Xavier Perrot and Andrew J. Oxenham. 2006. "Influence of musical and psychoacoustical training on pitch discrimination." *Hearing research,* **219**(1-2), 36-47.

[20] Mithlaj Jumana. 2018. "Effect of carnatic vocal musical training on temporal resolution, frequency discrimination and speech perception in noise in children." Masters Dissertation: *All India Institute of Speech and Hearing, Mysuru.*

[21] Mizener Charlotte P. 2008. "Enhancing language skills through music." *General Music Today,* **21**(2), 11-17.

[22] Moore David R., Melanie A. Ferguson, Lorna F Halliday and Alison Riley. 2008. "Frequency discrimination in children: Perception, learning and attention." *Hearing Research,* **238**(1-2), 147-154.

[23] Prithvi T. 2018. "Gap detection test using MLP toolbox-Development of normative in children (7-9 years)." Masters Dissertation: *All India Institute of Speech and Hearing.*

[24] Rauscher Frances, Gordon Shaw, Linda Levine, Eric Wright, Wendy Dennis and Robert Newcomb. 1997. "Music training causes long-term enhancement of preschool children's spatial-temporal reasoning." *Neurological Research,* **19**(1), 2-8.

[25] Ribeiro Ana Cláudia Mondini, Renata Coelho Scharlach and Maria Madalena Canina Pinheiro. 2015. "Assessment of temporal aspects in popular singers." *CoDAS.*

[26] Sangamanatha A.V., Jacqueline Fernandes, Jayashree Bhat, Manav Srivastava and S.U. Prakrithi. 2012. "Temporal resolution in individuals with and without musical training." *Journal of Indian Speech Language and Hearing Association,* **26**(3), 27-35.

[27] Saunders T. Clark and John M. Holahan. 1997. "Criteria-specific rating scales in the evaluation of high school instrumental performance." *Journal of Research in Music Education,* **45**(2), 259-272.

[28] Strait Dana L., Nina Kraus, Erika Skoe and Richard Ashley. 2009. "Musical experience and neural efficiency-effects of training on subcortical processing of vocal expressions of emotion." *European Journal of Neuroscience,* **29**(3), 661-668.

[29] Sutcliffe Paul and Dorothy Bishop. 2005. "Psychophysical design influences frequency discrimination performance in young children." *Journal of Experimental Child Psychology,* **91**(3), 249-270.

[30] Wong Patrick, Erika Skoe, Nicole M. Russo, Tasha Dees and Nina Kraus. 2007. "Musical experience shapes human brainstem encoding of linguistic pitch patterns." *Nature Neuroscience,* **10**(4), 420-422.

The Journal of Acoustical Society of India : Vol. 50, No. 3-4, 2023 (pp. 156-161)

A nonlinear acoustical exploration with Tagore verses to characterize readings, recitations and songs: A Case of Phase Transition?

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

[1]*Sir C.V. Raman Centre for Physics and Music, Jadavpur University*
[2]*Department of Humanities and Social Sciences, IIT Kharagpur*
[3]*Shrutinandan School of Music, Kolkata*
[4]*Department of Physics, Jadavpur University*
[5]*School of Languages and Linguistics, Jadavpur University*
e-mail: archibanerjee7@gmail.com

[Received: 14-03-2022; Revised: 21-05-2022; Accepted: 10-06-2022]

ABSTRACT

The work explores the variation of scaling exponents in a song, recitation and reading with the same lyrical content. Detrended Fluctuation Analysis (DFA) have been employed to find the long range temporal correlations (or the Hurst Exponent) present in each form of the auditory signal. Perceptually, it is known that the addition of rhythm and pitch, amplitude modulation distinguish between these different forms of audio signals, but the mathematical analogue of the same is still unknown. In this work, recordings were taken for 2 artists (1 male, 1 female) who were asked to read, recite and sing the entire lyrics of two (2) self chosen Tagore songs, each of which were later put to analysis. The rationale behind choosing Tagore's works is because of its strong lexical content. It was seen that Hurst Exponent is found to be the lowest in case of reading while it maximizes in case of song implying that the amount of long range correlation increases consistently with addition of rhythmic content and pitch, amplitude modulation in the audio signals. With this work, we tried to establish a critical value of Hurst exponent, above/below which transformation occurs to other forms analogous to critical temperature in phase transition of matter.

1. INTRODUCTION

A question has been whirling in the musical fraternity for generations - "Which impacts more? Melody or lyrics?"; the answer to which is still unknown. It is well established that both lyrics and melody play distinctly different but important roles in portraying the emotional content of a particular song. The audience listens to the lyric and melody together in the song and tends to get an idea about the mood of the song. But what would happen if the lyric part is separated from the melody and conveyed to the audience as a separate entity altogether. Would the emotional content of the song remain the same or the meaning would change altogether? Here in comes another aspect of reading, which is called recitation

(List, 1963). A recitation is a way of combining the words together so that they have a sense of rhythm and pitch / amplitude modulation, to convey the emotional content imbibed within. There exist other types of human sound communication besides speech, recitation and song. Among these are various signaling devices, as for example, drum, horn, and xylophone speech, and whistled speech. In addition to these forms there is instrumental music, hummed melodies, and imitations of animal cries. Speech, recitation and song display certain characteristics which, taken as a group, set them apart from the other forms stated above. They are (1) vocally produced, (2) linguistically meaningful and (3) melodic (Chow & Brown, 2018). None of the other forms of human sound communication that have been mentioned share all three of these traits. The first two characteristics of speech, recitation and song, vocal production and linguistic meaningfulness, could undoubtedly be utilized in developing a classification system. But, then the classification system would be very much dependent on the language used. In this study we envisaged to develop a scientific classification system taking into consideration song and verses / poem of the great Indian poet and Nobel Laureate Rabindranath Tagore. The rationale behind taking Tagore's works is its strong lexical content. In general it is difficult to separate his music from his literature, many of which went on to become lyrics for his songs later. Tagore's songs span the entire spectra of human emotions, depicted in his vast creative works ranging from *Brahmo* devotional hymns to purely romantic compositions. Two (1 male and 1 female) participants were asked to read, recite and sing the entire lyrics of two (2) self chosen songs, each of which were later put to analysis. The readings, recitations as well as the songs were analyzed with the help of a latest non linear technique called Detrended Fluctuation Analysis (DFA).

Speech and music signals shows a complex behavior (Bigrelle & Iost, 2000): at every instant components (in micro and macro scale: pitch, timbre, accent, duration, phrase, melody etc) are closely linked to each other (Pickover & Khorasani, 1986). These properties are peculiar of systems with chaotic, self organized, and generally, non linear behavior (Sanyal *et al.*, 2016). Therefore, the analysis of music using linear and deterministic frameworks seems not to be useful. So, a non-deterministic / chaotic approach is needed in understanding the speech / music signals. The Detrended Fluctuation Analysis (DFA) technique essentially computes the Long range temporal correlations (LRTC) present in the audio signals and uses a scaling exponent (called Hurst Exponent) to quantify them (Peng *et al.*, 1994). The parts having the exact lyrical content in the song as well as in the recital and reading part were extracted from the complete signal and analyzed with the help of DFA technique. From this analysis we intend to establish critical values, above which there will be transition from reading to recitation and consequently to a song. The case may be similar to the conventional phase transition, wherein the measurement of external condition at which the transformation occurs (generally, temperature in case of matter) is called phase transition.

2. EXPERIMENTAL DETAILS

Choice of audio signals : Two participants, one male and one female were asked to read, recite and sing the complete lyrics of two self chosen Tagore songs. The male participant selected two songs: "Maa ki tui porer dware pathabi tor ghorer chhele" and "Akash bhora surjo tara" whereas the female participant selected the following two songs: "Kolahol to baron holo" and "Jokhon esechhile ondhokare chand otheni". All the signals were normalized to 0dB. Each of these sound signals was digitized at the sample rate of 44.1 KHZ, 16 bit resolution and in a mono channel. Each of the 12 audio signals was then divided into 4 parts according to the sthayi, antara, sanchari and abhog part of the lyrics.

3. METHODOLOGY

DFA has been developed for quantifying correlation properties in non-stationary signals e.g., in physiological time series, because long-range correlations can also come from the artifacts of the time series data. In our study, DFA technique has been applied to quantify the scaling behavior of the fluctuations in each of the preprocessed audio signal parts and the scaling exponents were compared for each group of reading, recitation and song signals containing the exactly same lyrical content.

Detrended Fluctuation Analysis : To compute the Hurst exponent of a time series: $x_1, x_2, \ldots, x_N$ using Detrended Fluctuation Analysis or DFA technique (Peng *et al.*, 1994), firstly integration of x is done to form a new series $y = y_1, \ldots, y_N$, where

$$y(k) = \sum_{i=1}^{k} (x_i - \square x) \tag{1}$$

$\square x$ is the mean of $x_1, x_2, \ldots, x_N$.

The integrated series is then sliced into boxes of equal length or intervals of size n. In each box of length n, a least-squares line is fit to the data. The least square line represents the trend in that box. The coordinates of the straight line segments are denoted *by*$y_n(k)$. The root-mean-square fluctuation of the integrated series is calculated by

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N} (y(k) - y_n(k))^2} \tag{2}$$

where the part $y(k)–y_n(k)$ is called detrending. The relationship between the detrended series and interval lengths can be expressed as

$$F(n) \propto n^{\alpha} \tag{3}$$

where α is expressed as the slope of a double logarithmic plot log *F(n)* versus *log(n)* (as shown in representative Fig. 1 (a-c) . The parameter ? (scaling exponent, autocorrelation exponent, self-similarity parameter etc.) represents the auto-correlation properties of the signal.

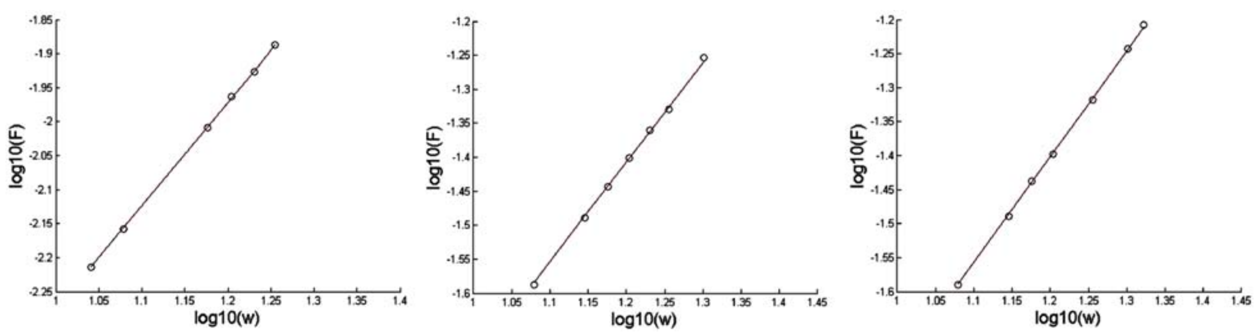


Fig. 1(a-c). DFA scaling exponent from double logarithmic plots for (a) read, (b) recite and (c) song.

DFA technique was applied following the NBT algorithm used in Hardstone *et al.*, (2012). The scaling exponent provides a quantitative measure of long range temporal correlation (LRTC) that exists in the audio signals. When the auditory waveform is completely uncorrelated (Gaussian or non-Gaussian probability distribution), the calculation of the scaling exponent results 0.5, also called white noise. When computing the scaling of the signal profile, the resulting scaling exponent, α, is an estimation of H. If α is between 0 and 1, then x was produced by a stationary process which can be modeled as a fGn process with $H = \alpha$. If α is between 1 and 2 then x was produced by a non-stationary process, and $H = \alpha – 1$ (Alvarez-Ramirez *et al.*, 2009).

4. RESULTS AND DISCUSSION

Comparing the DFA scaling exponent values for different parts of the renditions while reading, recitation and singing of the same lyrical content we observe the following trend (Fig. 2 and 3) for both of the artists. It is interesting to note that in most of the experimental cases, the value of scaling exponent is greater than 1, indicating that power law correlations exist in these audio signals originating from a non-stationary process.

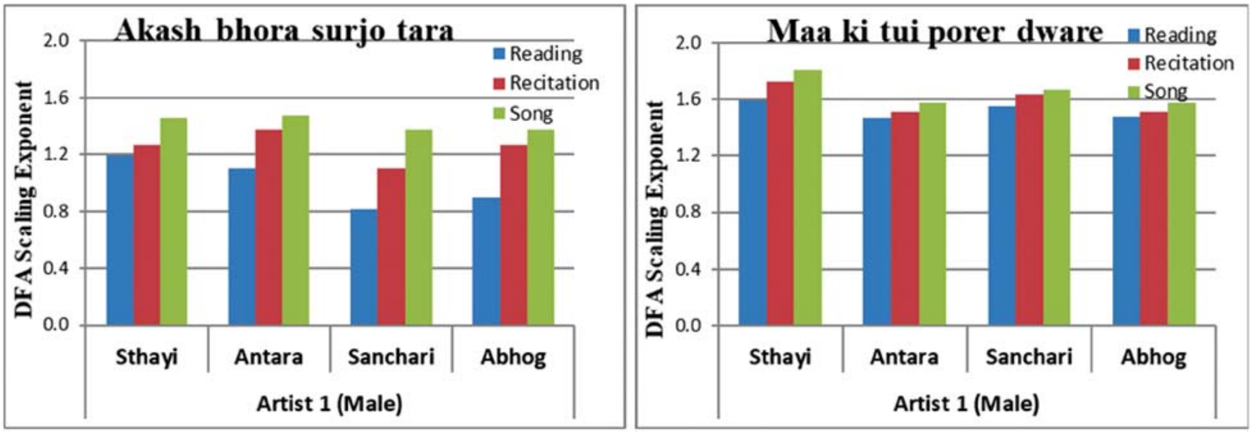


Fig. 2(a,b). Variation in DFA scaling exponent during reading, recitation and song for Artist 1 (male) in two self chosen songs.

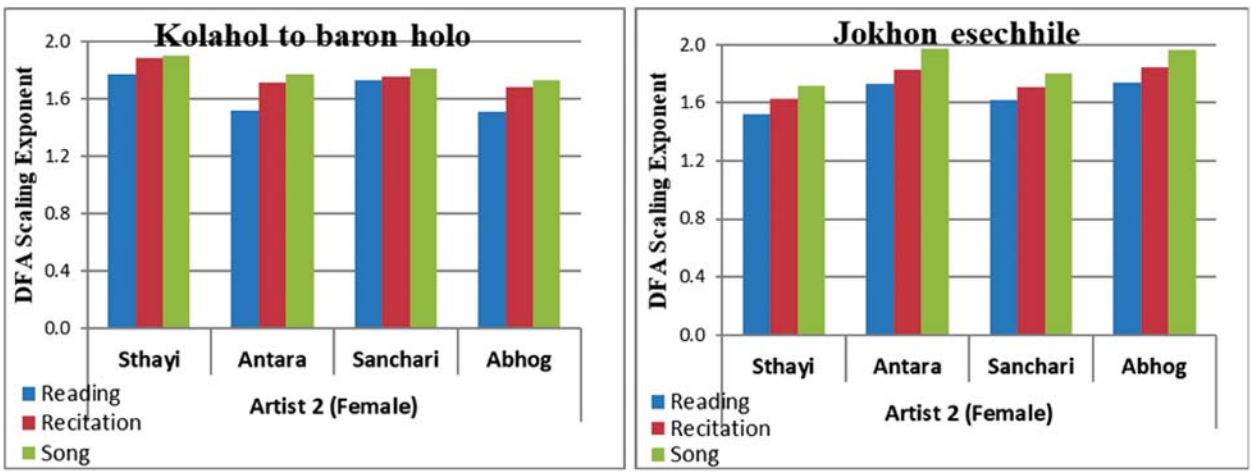


Fig. 3(a,b). Variation in DFA scaling exponent during reading, recitation and song for Artist 2 (female) in two self chosen songs.

A look at fig. 2 and 3 reveals that the DFA scaling exponent for the female artist is in general higher compared to the male artist indicating that the temporal correlations in the voice of the female participant are higher than the participant male voice. This can be attributed to the difference in source vocal characteristics of the two artists chosen.

From fig. 2a and b, it is evident that the scaling exponent is lowest in the "reading" part, increases in the "recitation" part and maximizes in the "song" part for all the four sections. The basic difference between these three forms of narration is caused mainly by the addition of rhythmic content, as well as, pitch and amplitude modulation which can directly contribute to the increase in the order of power law temporal correlations present in the auditory waveforms. In the recitation format, higher dependence on rhythmic content and amplitude modulations is expected than the reading format; whereas in the song format, further addition of pitch modulations is observed in general. This is the probable reason behind the gradual increment of scaling exponent from "reading" to "recitation" to "singing".

From fig. 2 and 3, we observe that for all the chosen 4 songs, the average increment of scaling exponent from reading to recitation is much higher than the increment from recitation to song. This again strongly points towards the dependence of temporal correlation on the rhythmic content and amplitude modulation present in the audio signal.

Further, in 3 of the chosen 4 verses, we see that both *antara* and *abhog* sections report very similar values of scaling exponent (more prominently so in the "song" versions). This could have happened due to the similarity in melodic and rhythmic pattern of *antara* and *abhog* sections, which is a commonly observed feature of Tagore songs.

Moreover, in fig. 2a, it is seen that the scaling exponent is greater in the *sthayi* and *antara* part while it decreases in the *sanchari* and *abhog* section. In case of fig. 2b though, *sthayi* and *sanchari* section have higher values of scaling exponent compared to the other two sections. This implies with the change of lexical content also, the scaling exponent varies significantly. Thus, the correlations present in any auditory time series is a manifestation of all the basic characteristics of sound that we know of and thus, highly dependent on both lyrics and melody present in the auditory waveforms. This is evident from the variation in scaling exponent even among different parts of the "reading" version of the verses chosen.

From fig.2 and 3, it is seen that the change in scaling exponent from one form of narration to other is highest in Sample 1 (Akash bhora surjo tara) by Artist 1 and lowest in Sample 2 (Maa ki tui porer dware) by Artist 1. This could have been caused by lesser modulation from one form to another by Artist 1 in Sample 2 as compared to Sample 1. Artist 2 (female) features similar kind of modulations in both of her chosen songs. Nevertheless, the values of scaling exponent for Artist 2 follows the same trend as in the case of Artist 1, with "reading" having the lowest scaling exponent and "song" having the highest scaling exponent. While *sthayi* section has the highest amount of temporal correlations present in fig. 3a, *antara* and *abhog* section have the higher values in fig. 3b.

5. CONCLUSION

In this study, we propose a novel classification algorithm with which three forms of narration i.e., plain reading, poetic recitation and singing can be distinguished mathematically from the acoustic waveforms with the help of a unique scaling exponent called the Hurst exponent. The main findings of this study can be summarized as under:

1. The scaling exponent, which is a manifestation of the long range temporal correlations present in the signal, increases uniformly for all the experimental pieces as one goes from "reading" to "recitation" and further to "song" keeping the lyrical content exactly same. The increase in self similarity of the signals can be attributed to the contribution of rhythm, amplitude and pitch modulation to the lexical content.
2. The samples were analyzed segregating the four sections of a literary piece i.e. sthayi and antara sanchari and abhog. It has been found that the scaling exponent significantly differs from section to section of different literary piece (even in the "reading" versions), indicating lyrics also have an effect on the amount of power law correlations in the signal. In fact, the variation in scaling exponent depends both on the melodic and lyrical content of the renditions. This is evident from the similarity of scaling exponent values in the antara and abhog sections of the "song" versions.
3. The difference between "reading" and "recitation" in terms of long range temporal correlations is much higher than the difference between "recitation" and "singing".
4. It is observed that the scaling exponent is in general higher for female voice compared to a male voice for the chosen two participants.
5. We have tried to find a critical value above which the transition occurs from one form of narration to another. The results of our experiment indicate that this criticality is dependent on the lyrical and melodic content of the verses chosen.

Analysis with a greater number of samples as well as participants will lead to the determination of a more accurate value of the critical point where the transition occurs from one form of narration to another.

The individual role of lyrics and melody in a song (as well as in recitation) can also be evaluated from this kind of study. This study is a precursor in that direction.

6. ACKNOWLEDGEMENT

AB acknowledges the Department of Science and Technology (DST), Govt. of India for providing (SR/CSRI/PDF-34/2018) the DST CSRI Post Doctoral Fellowship to pursue this research work. SS acknowledges the JU RUSA 2.0 Post Doctoral Fellowship (R-11/557/19) and Acoustical Society of America (ASA) to pursue this research.

7. REFERENCES

[1] Alvarez-Ramirez J., Rodriguez E. and Echeverria J.C., 2009. A DFA approach for assessing asymmetric correlations. Physica A: *Statistical Mechanics and its Applications,* **388**(12), 2263-2270.

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

[3] Chow I. and Brown S., 2018. A Musical Approach to Speech Melody. *Frontiers in psychology,* **9,** 247.

[4] Hardstone R., Poil S.S., Schiavone G., Jansen R., Nikulin V.V., Mansvelder H.D. and Linkenkaer-Hansen K., 2012. Detrended fluctuation analysis: a scale-free view on neuronal oscillations. *Frontiers in physiology,* **3,** 450.

[5] List G., 1963. The boundaries of speech and song. *Ethnomusicology,* **7**(1), 1-16.

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

[7] Pickover C.A. and Khorasani A.L., 1986. Fractal characterization of speech waveform graphs. *Computers & graphics,* **10**(1), 51-61.

[8] Sanyal S., Banerjee A., Patranabis A., Banerjee K., Sengupta R. and Ghosh D., 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.

The Journal of Acoustical Society of India : Vol. 50, No. 3-4, 2023 (pp. 162-168)

Study of long term effect (LTE) of Sudarshan Kriya (SK) and music listening (ML) using EEG Analysis of Human Brain Response (HBR)

Rana Bishwamitra and Maringanti Himabindu*
Department of Computer Applications,North Orissa University,Odisha-757 001, India
e-mail: profhbnou2012@gmail.com

[Received: 29-03-2022; Revised: 11-05-2022; Accepted: 13-06-2022]

ABSTRACT

Stress is the major factor for numerous mental and physical health disorders. Meditation brings positive changes at psychology and physiology level which helps to reduce the stress. As per the existing research, meditation helps to decrease the anxiety, stress and increase in relaxation, focus and produce other well-being related to psychological level. These changes can be understood by studying the neurophysiological effects of meditation on brainwave signals using EEG (Electroencephalography). The aim of this paper is to review the studies based on neural mechanics using EEG brainwave signals of various meditation practices. The objective of the present work is to study the long term effects of Sudarshan Kriya, a rhythmic breathing technique founded by Sri Sri Ravi Shankar of Art of Living, on the human brain, by studying the response of the brain through Electroencephalogram (EEG) analysis. Hence EEG recording of experts (Who have practiced SK for around 10-15 years) was done. The spectral waveform resulting out of these recordings are analyzed mathematically and programmatically using known string matching algorithm. As a result of this matching, logic patterns of commonality that way be repeated across the spectrum are identi-fied / mined / extracted. Thus the patterns form BLOC (Base level of consciousness) components. The patterns so obtained can now be interpreted and standardized as the effect of Sudarshan Kriya (SK) practice for a long time. The response of EEG so obtained is compared with EEG obtained from music listening experi-ences which have conventionally used as a source of meditation using the same analytics. So analysis of EEG spectra after sufficiently long practice of SK and music listening and its effect on the brain waves (α, β, δ, □) during a period of 15 minutes is the objective of the present work.

1. INTRODUCTION

Meditation and EEG : Meditation has positive effects on human psychology and at physiological level such as reduction in stress, enhancement in executive functions of brain, boosting immunity and various parameters pertaining to well-being, both physical and mental. Human brain consists of millions of neurons which plays an important role for controlling behaviour of human body with respect to internal / external motor / sensory stimuli. These neurons will act as information carriers between human body and brain. Efficient understanding of the cognitive behaviour of humans can be done by analyzing either signals or images from the brain. Human behaviour can be visualized in terms of motor and sensory states such as, eye movement, lip movement, remembrance, attention, hand clenching and other gestures etc. These states are related with specific signal frequency which helps to understand functional behavior of complex brain structure. Electroencephalography (EEG) is an efficient modality recorded from the scalp surface area, which helps to acquire brain signals that correspond to various conscious states. These signals are generally categorized as delta, theta, alpha, beta and gamma based on signal frequencies ranges from 0.1 Hz to more than 100 Hz. In EEG, four types of brain waves exist classified according to the frequency range, *viz.*, Alpha(α) in the frequency range of 8-13.9 Hz , Beta (β) in the frequency range of 14-30 Hz, Theta (θ) in the frequency range of 4-7.9 Hz and Delta (δ) in the frequency range of 1-3.9 Hz[1].

Consciousness level : Consciousness is a mental and awakened state of a human, which lets him / her be aware of the surroundings; which in turn helps in reacting to the external / environmental stimuli or acting in the environment. The various stages of consciousness include Super-consciousness (SU-C), Consciousness (CON), Semi-consciousness (SE-C) and Unconsciousness (UN-C). These stages are important to be analyzed and understood, for mental and psychological health; as the functions and abilities of man's mind in these various stages are varied and different, sometimes overlapping, with fuzzy boundaries called as Alpha states (Random Eye Movement) or SE-C states.

The level of consciousness (LOC) from which we receive guidance is the conscious state; the rational awareness that usually guides our daily decisions. When we receive input from the senses, they are analyzed and decisions are taken based on this information, using this conscious level of guidance. This process is also strongly affected / biased by the opinions of others, which can cloud our ability to draw true conclusions. Dividing and separating the world into either / or categories, the conscious level of awareness is problem-oriented. It is difficult to be completely certain of decisions drawn from this level, because the analytical mind can see all the possible solutions. But ultimately it doesn't have the ability to distinguish which one is best. Usually, experience derived from the unconscious mind and / or intuition helps in choosing the best solution[2]. The LOC could be esti-mated by using the measurement scale, called as the AVPU. The A is the measure of alertness and orientation, the V is the response to verbal stimuli, the P is the response to pain and the U is the measure of unresponsiveness. The second scale which is used by medical practitioners is the Glas-gow Coma Scale, which measures eye opening, verbal response and motor response and takes the sum of the scores to decide. The third scale is the PEARRL, which assesses the presence of pupils (P), their dilation equal (E), and (A), round in shape (R), regular in size (R) and response to light (L). The latest scale was suggested and copyrighted by Dr. David R.Hawkins[2,3] and so called as the Hawkin's Scale suggests the use of the Map of Consciousness. This map is a numeric scale over which the distinction between positive and negative, truth and false, power and force could be made. He believed that every thought or intention creates a morphogenetic field; a magnetic field and that these energy fields form a map consisting of nodes, that could easily be measured, by using Kinesiology. Later, Dr. John Diamond refined the scale to include Behavioral aspects, wherein the strength of muscles was measured when the patient is under emotional and intellectual stimuli. Dr. Hawkin's work is more recent, and so taken much farther to include detection of positive / negative stimuli, anabolic / catabolic and truth / false, calibrating statements, thoughts, photos, art and Music. His experiment constituted of putting a two-finger pressure on the wrist of the patient and checking his ability to withstand the pressure. Such a technique is probably being used by the Lie-Detectors as part of the Forensic Polygraphic Testing[9,10,11].

Music, a universal human phenomenon is structured tonal sound moving in time and space. From its origins in the primitive imitation of nature's sounds like songs of birds, calls of animals, rustling of leaves and waves of the ocean, music has evolved into organized forms that are varied in style and idiom from century to century and culture to culture. Music therapy interventions can be designed to Promote Wellness, Manage Stress, Alleviate Pain, Express Feelings, Enhance Memory, Improve Communication, Promote Physical Rehabilitation[3,4]. Tanpura is a multistringed instrument extensively used as a drone instrument and is an integral part of classical music in India. The resounding twangs of the strings create the atmosphere of Indian classical music. Its sound is considered very sweet and melodious, and it stimulates both the musician and the audience. Drone can provide contrast but is not prompting a response.

The drone environment is free of semantic con-tent, such as melody or rhythm, similar to an acoustical Ganzfeld[5], where no delimitable (sound) objects can be grasped or (re)cognized. In the Ganzfeld, cognitions arise spontaneously out of in-trinsic activity[6]. A number of earlier studies try to decipher the effect of tanpura drone on human brain using different techniques[7-8]. Thus, it will be quite interesting to compare the brain re-sponse in a meditative state to that of tanpura drone induced state.

In this work the main aim is to compare the response of human brain under the effect of sim-ple meditation and to compare that with the effect of simple acoustical stimuli - tanpura drone which is long known to have meditative features. For this, we took EEG of 3 participants who are experts in Sudarshan Kriya Yoga, when they performed the meditation as well as the EEG of the same 3 participants when they listened to tanpura drone stimulus. The delta and theta band power of the two frontal electrodes F3 and F4 were computed for all the participants during meditative as well as music induced states and compared. The results show interesting new features which can be further utilized for general wellbeing of common people.

2. MATERIALS AND METHODS

Electroencephalogram (EEG) recording of experts (Who have practiced SK for around 10-15 years) was done. The spectral waveform resulting out of these recordings are analyzed mathematically and programmatically using known string matching algorithm. As a result of this matching logic, patterns of commonality that way be repeated across the spectrum are identified / mined / extracted. Thus the patterns form BLOC(Base level of consciousness) components. The patterns so obtained can now be interpreted and standardized as the effect of SK practice for a long time. So analysis of EEG spectra after sufficiently long practice of SK and its effect on the brain waves (α, β, δ, □) during a period of 15 minutes is the objective of the present dissertation work.

3. EXPERIMENTAL WORK

Power Spectral Intensity (PSI) : In order to eliminate all frequencies outside the range of interest, data was band pass filtered with a 0.5-35 Hz FIR filter. The amplitude envelope of the delta (0.5-4 Hz) and theta (4-8 Hz) frequency range was ob-tained using wavelet transform technique proposed by Linkenkaer-Hansen *et al.* (2001). The amplitude enve-lope of the different frequency rhythms were obtained for 'before meditation', 'with meditation' as well as 'without meditation' conditions for the two frontal electrodes (F3, F4). A number of studies validated the importance of frontal electrodes in case of cognitive processing[13,14]. So, we chose to study the variation of scaling exponent corresponding to various frequency rhythms in the five frontal electrodes while listening to meditation of contrast emotions.

To the time series data $[x_1, x_2, \ldots, x_N]$, we perform the Fast Fourier Transform (FFT) using fft function in MATLAB and the result obtained is denoted as $[X_1, X_2, \ldots, X_N]$. A continuous frequency band from f_{low} to f_{up} is sliced into K bins, which can be of equal width or not. The bins used are δ (0.5-4 Hz), θ (4-7 Hz), α (8-12Hz), β (12-16 Hz), and γ (16-30 Hz). For these bins, we have band = [0.5, 4, 7, 12, 30, 50]. The Power Spectral Intensity (PSI) of the k^{th} bin is evaluated as given in Bao *et al.* (2011).

$$PSI_k = \sum_{i=\lfloor N(f_k/f_s)\rfloor}^{\lfloor N(f_{k+1}/f_s)\rfloor} |Xi|, \; k = 1,\; 2, \ldots,\; K-1 \tag{1}$$

Where f_s is the sampling rate, and N is the series length. The delta and theta power values have been computed using the above algorithm for the five frontal electrodes corresponding to various experimental conditions. Each window is converted into the frequency domain using Fast Fourier transform (FFT). The average power corresponding to each experimental condition was computed for all the frontal electrodes.

Subject Summary : The three (3) subjects were selected such that their responses would be useful to us for extraction of benchmark data, which a novice should achieve after a significant practice of Sudarshan Kriya (SK). These were all AOL teachers with more than 10 years of practice of SK. For JU47, the re-cording of EEG started after he had a session of about 20 minutes of meditation, while for others EEG was taken for pre-meditative state also.

Experimental Protocol : The EEG experiments were conducted in the afternoon (around 2 PM) in an air conditioned room with the subjects sitting in a comfortable chair in a normal diet condition. All experiments were performed as per the guidelines of the Institutional Ethics Committee of Jadavpur University. All the subjects were prepared with an EEG recording cap with 19 electrodes (Ag/AgCl sintered ring electrodes) placed in the international 10/20 system. Impedances were checked below 5 k Ohms. The EEG recording system (Recorders and Medicare Systems) was operated at 256 samples/s recording on customized software of RMS. Raw EEG signals were filtered using a low and high pass filter with cut-off frequencies of 0.5 to 35 Hz. The electrical interference noise (50 Hz) was eliminated using notch filter. Muscle artifacts were removed by selecting the EMG filter. The ear elec-trodes A1 and A2 linked together have been used as the reference electrodes. The same reference electrode is used for all the channels. The forehead electrode, FPz has been used as the ground electrode. Each subject was seated comfortably in a relaxed condition in a chair in a shielded measurement cabin. After initialization, a 30 min recording period was started, and the following protocol was followed: First 5 min Control (Pre-Meditative stage) Second 15 min (Meditative stage) Finally 5 min Rest (Post-Meditative stage). After initialization, the following protocol was observed for the three subjects:

Subject Id: JU46 Part 1: Rest state: 00:00:30 - 00:05:00
Part 2: Meditation State: 00:05:10: - 00:20:00
Part 3: After Meditation: 00:20:30 - 00:25:30

Subject Id: JU47 Part 1: Meditation State: 00:00:30 - 00:15:00
Part 2: Post Meditation State: 00:15:10: - 00:20:00

Subject Id: JU48 Part 1: Rest state: 00:00:30 - 00:05:00
Part 2: Meditation State: 00:05:10: - 00:20:00
Part 3: After Meditation: 00:20:30 - 00:25:30

The same protocol was repeated with tanpura drone as the input music replacing the meditative state.

Response from F3 and F4 electrodes was analyzed and the following results obtained. Data analysis and interpretation between Delta and Theta of subjects ID: JU46, subjects ID: JU47 and subjects ID: JU48 by following F3 and F4 electrodes for observation for state of brain.

4. RESULTS AND DISCUSSION

The EEG response for the 3 individual participants in response to meditation and again in response to tanpura drone or the music clip have been illustrated in Fig. 1-3 respectively. The delta and theta power values have been taken into consideration, since these play the most important role in higher order cognitive and thinking skills. Fig. 1 shows the delta and theta response for the 1st participant in the three experimental conditions for meditation as well as music induced states.

As is seen from the figures the delta frequency region is the most affected during the meditative state in the left frontal electrode, while the right frontal electrode shows a predominance of theta waves. The arousal in delta waves is much more pronounced in case of meditation compared to music in F3, while in music, we see significant elicitation in delta waves after the removal of music stimulus. In F4, however the response is much stronger under musical stimuli.

In this case, however we see that both the delta and theta power increases to a much larger extent under the effect of musical stimuli in both F3 and F4 electrode. The delta power increases signifi-cantly in the post meditative state as well as in the post music stimuli induced state in both the hemispheres. The theta power again predominates in the music stimuli induced state corresponding to the right frontal electrode and consistently increases during the period of musical stimulus.

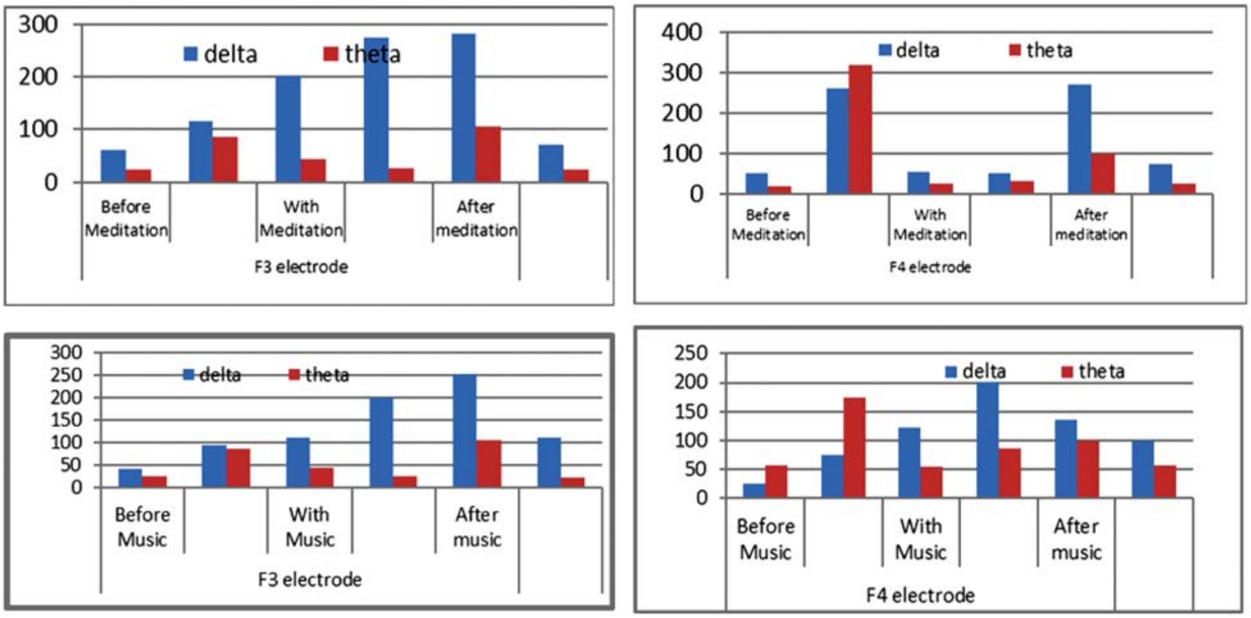


Fig. 1. Delta and theta power variation for meditation and music.

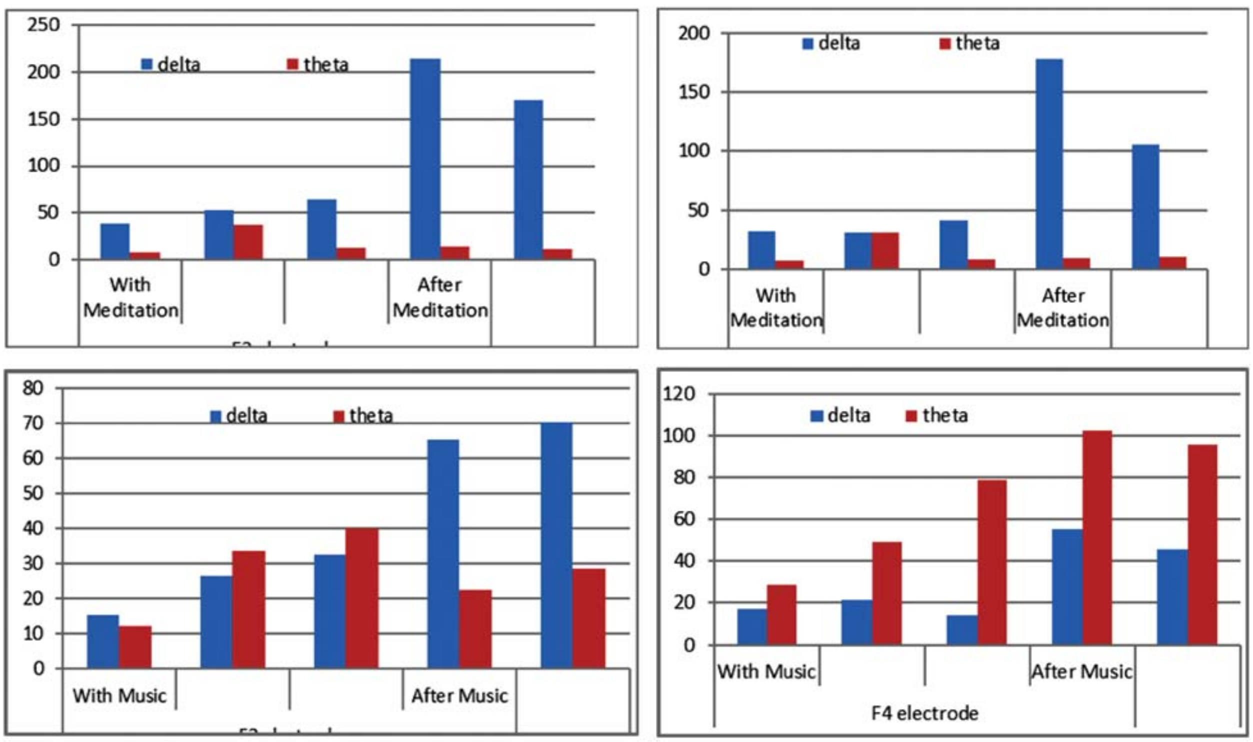


Fig. 2. Delta and theta power variation for meditation and music JU 47.

In this case, we find comparable results for both music and meditation data. While the delta power increases significantly under the effect of meditation, both delta and theta power increases for music in both the F3 and F4 electrode. In the post-meditative state, for both the electrodes we find the delta power remains on the higher side while the same decreases in the post music induced state.

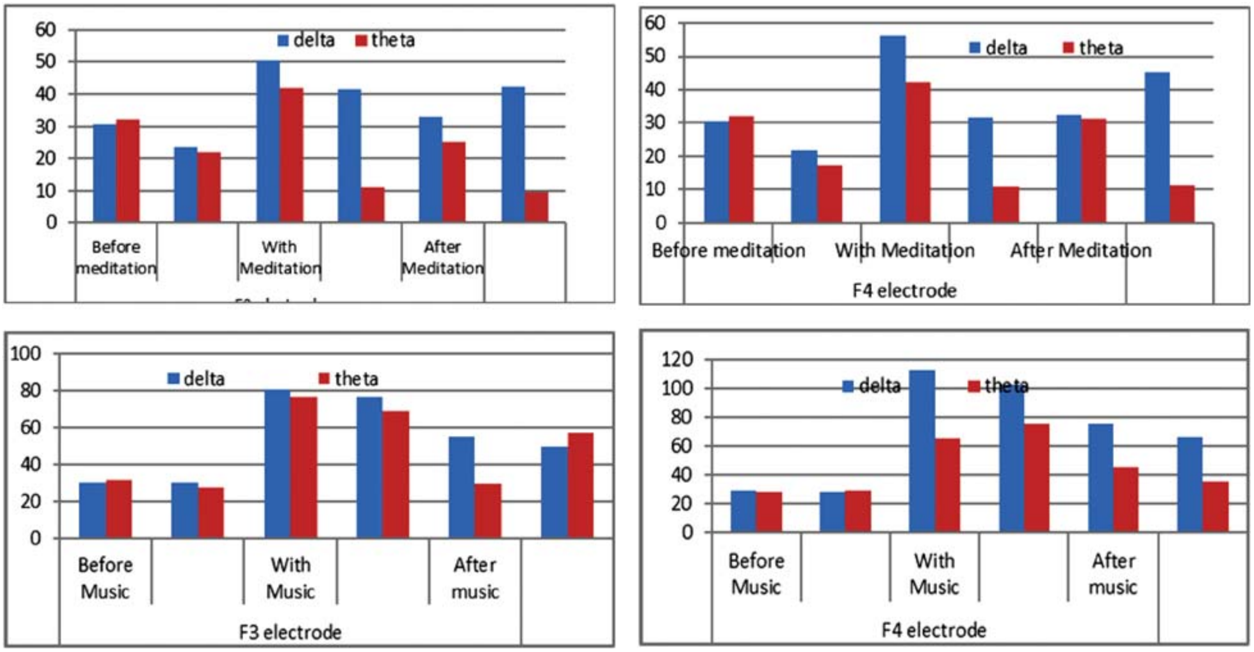


Fig. 3. Delta and theta power variation for meditation and music JU 48.

5. CONCLUSION

The objective of the present paper is to study the effects of meditation, especially Sudarshan Kriya founded by Sri Sri Ravi Shankar on the conscious state of humans though the recording of EEG signals and analyzing them for excitation in the four levels of brain activation states, viz., alpha, beta, delta and gamma. Non-linear mathematical and signal processing techniques like MFDFA and MFDXA are used for extracting the common patterns of response, that would help us decide upon a baseline by conducting the experiment on experienced subjects. Three phases of EEG recording, viz., pre-meditation, mid-phase and post-meditation, are analyzed for the excitation and peak amplitude of the brain waves. Comparing the results of the meditation data, with the music induced EEG, we see that there is a need for person specific approach in this kind of study. The results for the meditative states is very specific to a particular person, but in general shows the predominance of delta waves, while in case of music, we find the predominance of theta waves. This study is not the end, but a beginning to much more in the field of cognitive neuroscience and psychology.

6. REFERENCES

[1] J. Satheesh Kumar and P. Bhuvaneswari, 2012. "Analysis of Electroencephalography (EEG) Signals and Its cCategorization - A Study" *International Conference on Modeling,Optimization and Computing, Procedia Engineering, (ICMOC),* **38,** 2525-2536.

[2] Hima Bindu M., 2015. "CONSCIOUSNESS AND MUSIC" *Proceedings of 22nd Internationational Symposium Frontiers of Research in Speech and Music (FRSM)*, November 23-24, 2015, (Indian Institute of Technology) IIT Kharagpur, pp. 123-127

[3] Saroj Kothari, 2015. Professor & Head, Department of Psychology Govt. M.L.B.P.G. Girls College, Indore," Effects of Dance and Music Therapy", *International Journal of Research - Granthaalayah Innovation in Music & Dance*, ISSN- 2350-0530(O), pp. 2394-3629.

[4] Scartelli J.P., 1982. The effect of sedative music on electromyography biofeedback assisted relaxation training of spastic cerebral palsied adults. *Journal of Music Therapy*, **19,** 210-218.

[5] Metzger, W., 1930. Optische Untersuchungen am Ganzfeld. *Psychological Research*, 10.1007/BF00406757, **13,** 6-29.

[6] P¨utz P., Braeunig M. and Wackermann J., 2006. EEG correlates of multimodal Ganzfeld induced hallucinatory imagery. *International Journal of Psychophysiology*, **61**(2), 167-178.

[7] Maity A.K., Pratihar R., Mitra A., Dey S., Agrawal V., Sanyal S. and Ghosh D., 2015. Multifractal detrended fluctuation analysis of alpha and theta EEG rhythms with musical stimuli. Chaos, *Solitons & Fractals,* **81,** 52-67.

[8] Sanyal S., Nag S., Banerjee A., Sengupta R. and Ghosh D., 2018. Music of brain and music on brain: a novel EEG sonification approach. *Cognitive Neurodynamics,* pp. 1-19.

[9] Ali Rajabi Kheirabadi, Jamshid Tabeshpour and Reza Afshari, 2012. "Comparison of Three Consciousness Assessment Scales in Poisoned Patients and Recommendation of a New Scale: AVPU Plus", *Asia pacific Journal of Medical Toxicology*.

[10] Georgios MATIS and Theodossios BIRBILIS, 2008." The Glasgow Coma Scale - a brief review Past, present, future", *Acta Neurol. Belg.,* **108,** 75-89

[11] Carla M. Thompson, 2012. The Map of Consciousness- Hawkin's Scale.

[12] Zorick T. and Mandelkern M.A., 2013. Multifractal Detrended Fluctuation Analysis of Human EEG: Preliminary Inves-tigation and Comparison with the Wavelet Transform Modulus Maxima Technique. *PLoS ONE,* doi:10.1371/journal.pone.0068360, **8**(7), e68360.

[13] Banerjee A., Sanyal S., Patranabis A., Banerjee K., Guhathakurta T., Sengupta R. and Ghose P., 2016. Study on brain dynamics by non linear analysis of music induced EEG signals. *Physica A: Statistical Mechanics and its Applications,* **444,** 110-120.

[14] Sanyal S., Banerjee A., Pratihar R., Maity A.K., Dey S., Agrawal V. and Ghosh D., 2015. De-trended Fluctuation and Power Spectral Analysis of alpha and delta EEG brain rhythms to study music elicited emotion. *In Signal Processing, Computing and Control (ISPCC), International Conference on IEEE*, pp. 205-210.

[15] Bao F.S., Liu X. and Zhang C., 2011. PyEEG: an open source python module for EEG/MEG feature extrac-tion. *Computational intelligence and neuroscience.*

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)