Original Research Article Medical

International Journal of Pharma and Bio Sciences

ISSN
0975-6299

CREATION OF VOICE DATABASE, ACOUSTIC ANALYSIS AND STANDARDISATION OF NORMAL INDIAN VOICES

LATHADEVI H T*[1], MALIPATIL SR[2], S P GUGGARIGOUDAR[3]

*[1,2,3]*Department of Otolaryngology and H& N Surgery, BLDE University's Shri BM Patil Medical College & Hospital, Vijaypur, Karnataka, India..*

ABSTRACT

Acoustic analysis is used to assist differential diagnosis, documentation and evaluation of treatment for voice disorders. Clinical data have shown that jitter and simmer are indices of voice pathology or perceptual hoarseness but are more commonly used as an outcome measure.A voice with some periodicity can now be analyzed with a computerized acoustic analyzer, a relatively newer technique that can be widely used in the clinical practice.The purpose of the study was to create a normative Indian voice database and propose a standardization for normal acoustic parameters for Indian voices. 1000 Normal voice samples were collected from college students (male and female) aged 18-28 yrs. A sustained vowel /a/ was recorded and analyzed for parameters like Jitter, Shimmer, Harmonic noise ratio and Fundamental frequency (Fo) using a software. The Mean, SD, range of the voice parameters were calculated from the sample voices. The value of ranges of mean Jitter was 0.00 to 0.03635, mean Shimmerwas 0.06 to 0.2506 and Harmonic Noise Ratio was 8.31 to 30.73db.

KEYWORDS: ***Voice database, acoustic analysis, parameters of normal voice.***

LATHADEVI H T*

Department of Otolaryngology and H& N Surgery, BLDE University's Shri BM Patil Medical College & Hospital, Vijaypur, Karnataka, India.

Received on: 12-12-2016
Revised and Accepted on: 09-03-2017
DOI: http://dx.doi.org/10.22376/ijpbs.2017.8.2.b349-354

INTRODUCTION

Communication and expression through sound are found in most of the animals. However, it is a highly sophisticated and complex skill in humans. Spoken speech is the verbal communication that realizes the language in to sound output. This skill is acquired laboriously throughout the growth of the child into adulthood.Voice is the continuous sound produced by vocal cords that provides the basic sound to the organs of articulation which modulate it and finally produce speech. Essentially voice is the acoustic output of the vibrations of the vocal cords and the coloring of this output by the vocal tract.The sound of each individual's voice is his/her signature that is, entirely unique not only because of the actual shape and size of an individual's vocal cords but also due to the size and shape of the rest of that person's body, especially the vocal tract, and the manner in which the speech sounds are habitually formed and articulated. The voice gets affected when there is change in the architecture or the function of any of the above mentioned organs. An abnormal voice is variously described as hoarse, husky, breathy, harsh, rough etc. Moreover, our voices change throughout our lifetime, but there are also minor, and sometimes major, fluctuations throughout a day. Still, we are capable of differentiating normal voices from abnormal ones. But there is no universal, objective way of telling the exact difference between normal and abnormal voices. Since voice is essentially an acoustic output, it is possible to analyze its physical characteristics. It is logical to assume that our brain uses specific acoustic parameters to decide whether the voice is normal. The present study attempts to create a large database of normal voices, extract their physical parameters and statistically recognize the determinants of a normal voice.

Physiology of voice production

Voice is produced by vibration of vocal cords which produce the fundamental note. This is modified by the so called vocal tract which includes throat, nose and mouth. In fact, vocal tract is a series of acoustic filters which reinforce or decrease the original fundamental note.These filters are capable of assuming different shapes of varying sizesin oral and pharyngeal cavities. These changes in shape produce characteristic formants which distinguish one vowel from other. For example, the vowel [a:] have frequency bands in the region of 800 Hz and 1100 Hz.

Phonation

The Myoelastic- Aerodynamic theory or tonic theory (Van den Berg)[1] is generally regarded as the most widely accepted model to explain the mechanics of voice. According to this theory, when the vocal folds adduct in the midline due to action of interarytenoid muscle, there is build up of subglottic pressure. This causes the vocal fold to separate. When they adduct due to elastic recoil, the velocity of airflow increases in between vocal cords and the pressure between the vocal folds decrease (Bernoulli's principle). The decreased air pressure coupled with elastic recoil of vocal folds causes them to move back to midline. This is one cycle of vibration which repeats approximately 125 Hz in males and 225 Hz in females.This was questioned by Husson[2] who put forward his neuromuscular neurochronaxic or clonic theory of vocal fold vibration. They contended that every single vibration of the vocal cords was due to an impulse from the recurrent laryngeal nerves and the acoustic centre in the brain regulated the of vocal fold vibration. But this theory was disproved effectively. The problems with this theory was that theleft laryngeal nerve to vocal fold has a longer pathway/course than the right one; so nerve impulses should take longer than the right side; this means that vocal folds should vibrate out of phase which is not so.When tracheostomy is done the phonation does not occur even with effect the neural impulses - this means that air pressure is a component to phonation and not just neural impulses.Hence the voice requires a power source (lungs, abdominal muscles and back muscles), Oscillator (vocal folds) and resonator (vocal tract which includes oral cavity, oropharynx, supraglottis etc.)

MATERIALS AND METHODS

Collection of test subjects

Young healthy adults, between the age of 18 and 28 years, both male and female were selected for the study. All of them gave an informed consent. They underwent thorough clinical evaluation and those having any pathological condition or even suspicion of it were excluded. They were subjected to assessment of their voices by a speech therapist and an ENT surgeon. Only those who were certified as having normal voice were selected for the study. Finally there were 1000 tests subjects left.

Training and Recording of Voice Samples

Each person was first trained to produce a sustained vowel 'a' at comfortable loudness and pitch.

Recording

Recordings were made in a sound-treated chamber using a unidirectional microphone (Sony Audio-Technical 250XL) at a distance of 5cm in front of the lips. The sustained /a/ vowel signal was recorded for minimum of 3 seconds using the application PRAAT[3] version 5.4.04. The intensity was controlled using the VU meter built into PRAAT. The sustained vowel is preferred over regular speech in vocal acoustic assessment as it provides more reliable results.[4,5]

Sampling

A spectrograph of each sample was extracted and the most stable and uniform 1 sec slice was selected and saved as a sound file. A PRAAT script was developed that could batch process all files into a folder to extract four parameters i.e., Jitter, Shimmer, Harmonic Noise Ratio and Fundamental Frequency and out to an excel sheet. Each parameter was analyzed for different variations like mean, SD, range, etc. These ranges of parameters were compared with other normative data. All the parameters were determined with 95% confidence interval.

RESULTS

Table1
Voice parameters obtained from acoustic analysis

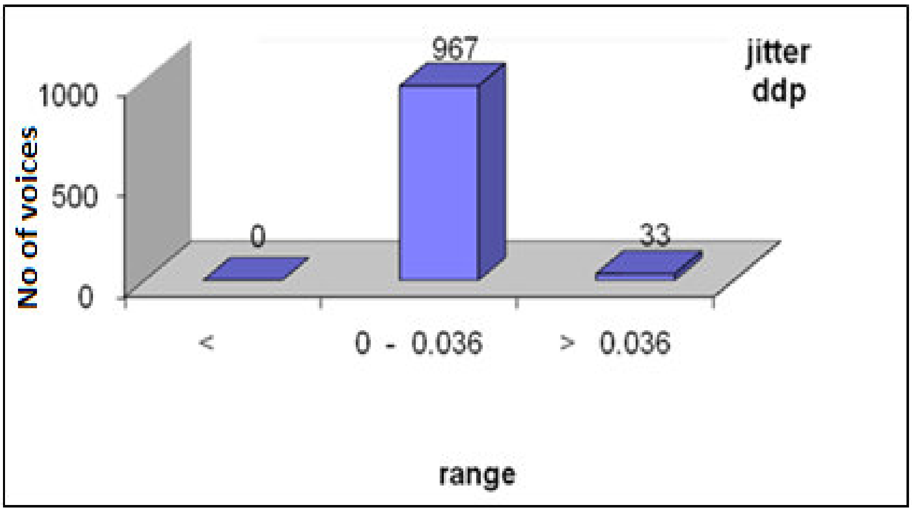
Parameter	Mean +/-1.96 SD 95%confidence interval
F0, fundamental frequency	201 to 928 hz
Jitter (ddp)	0 to 0.036352484
Shimmer(dda)	0.063955802 to 0.250624922
HNR (Harmonic noise ratio)	8.31738dB to 31.738Db

The Table.1 gives the total range of the values of the population under study. Especially the range of Fundamental Frequency is more. But the range of parameters like jitter, Shimmer and Harmonic noise ratio is almost same as other studies.

Table2
Studies showing acoustic measures for various authors

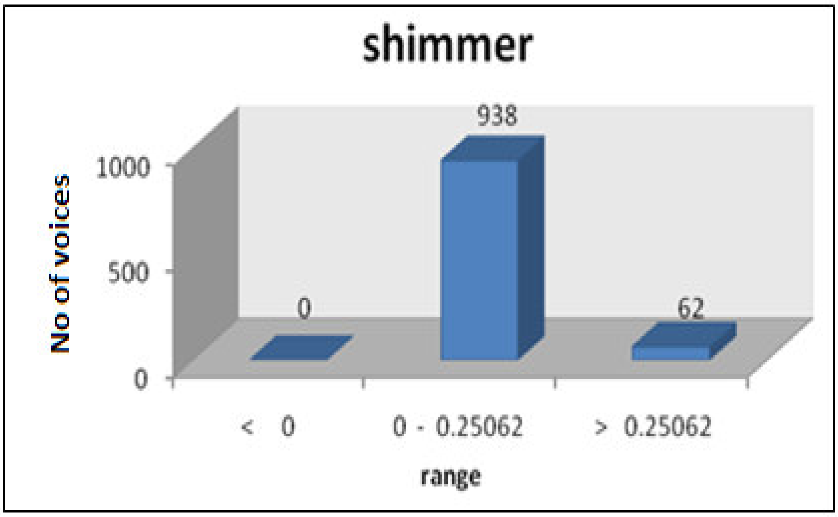
Sl/No.	Author	Jitter –male	Jitter female	Shimmer- male	Shimmer-Female	HNR-male dB	HNR-female Db	Fo-Male Hz	Fo- female Hz
1.	Williamson[7]	Less than 1.04%		Less than 3.81%		Less than 20		128	225
2,	Simone[8] N=80	0.37%	0-87%	63.77%	65.17%	1.06db	-1.64	127.6	215.45
3,	KC TORAN[9] n=50	0.14%	0.14%	1.6%	1,6%	25.81%	25.88%	170	246.45%
4.	ANC Fillippe[5] 20 m+20f	0.49	0.62	0.22	0.22	9.56	10.9	119	205
5.	CC Wang[10] 45m+45f	0.56	0.66					118.3	203.2
6.	K Aries[11] N=70	0.46%	0.87%	0.23%	2.72%	0.13	0.12	130.6	218.38
7.	Bonzi[12] N=72	0.36 Local%	0.31 Local%	3+/- 0.9	2.7+/-1.1	20+/-2	21+/-3		
9	Lathadevi et Al N=1000	0.018176242		0.1572899		20.024		120	220

Graph 1
Jitter range



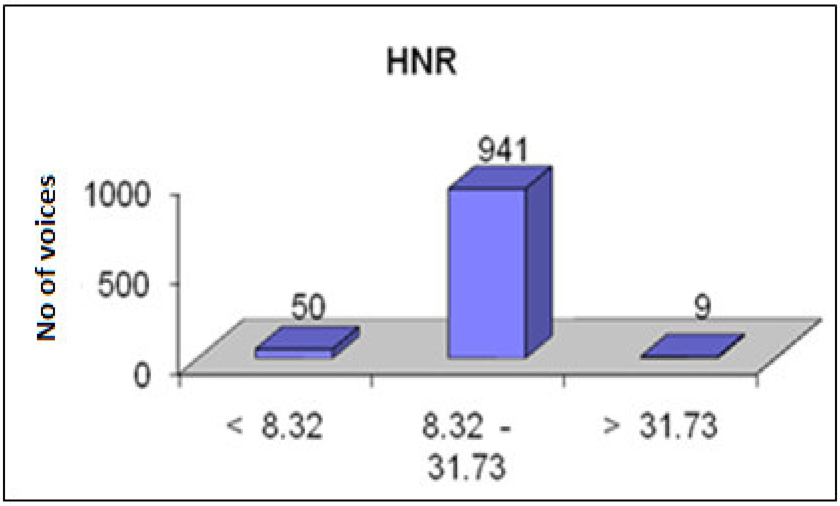
Graph.1 shows the number of voice population falling within the range of quoted range of jitter values. For e.g.The 967 voice samples have a range of 0.0 to 0.036Hz jitter.

Graph2
Shimmer range



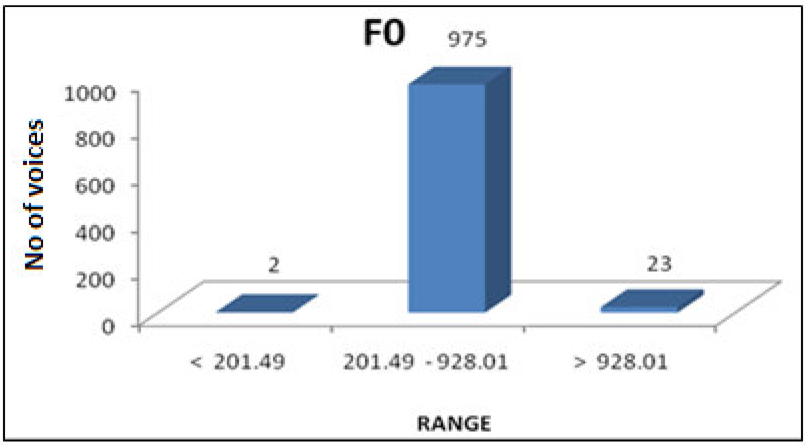
Graph.2 shows the number of voice population falling within the range of quoted range of shimmer values.The 938 voice samples have a range of 0.0 to 0.25062 shimmer.

Graph 3
HNR RATIO



Graph.3 shows the number of voice population falling within the range of quoted range of HNR values.The 941 voice samples have a range of 8.32 to 31.73dB.

Graph 4
Frequency Range F0



Graph.4 shows the number of voice population falling within the range of quoted range of Frequency values.

DISCUSSIONS

Voice assessment which was subjective so long has now been aided by objective analysis tools. This is essentially due to the advances made in precise sound capturing equipment and fast and accurate acoustic analysis software. Since the methods are simple and non-invasive they are being used widely throughout the world. A number of acoustic analysis software like Dr Speech, MDVP, PRAAT, and Vaghmi have been developed. Some of them come embedded in dedicated hardware. However, a simple set up can be established with a moderate computer, a good quality microphone and recording software. Some analytical software provides recording facility built into them. A sound treated room is mandatory to avoid extraneous noise and have a faithful recording. This arrangement provides us with extensive physical information about the test sound. The parameters so obtained for a particular sound can be assumed to be specific for that sound. Our hypothesis is that human brain perceives a voice as normal on specific parameters common to all normal voices. The present study aims at statistically finding out such parameters common to all test subjects having normal voice. These can later be used as a yardstick to diferentiate abnormal voices from the normal. There is limited literature available that addresses standardization of parameters of normal voice. Even in available data the number of normal voices was not more than 200[5,8,9,10,11,12]. In our study we have collected 1000 normal voices which would give more dependable results. This database can also serve other researchers as a database repository.From the clinical standpoint of view, assessment of voice signal has become noninvasive, less expensive and commonly available[6]. In our study we have used PRAAT, acoustic analysis software. This is being in a large number of laboratories throughout the world[7].PRAATis an efficient computer software package used for the analysis of speech in phonetics. It is developed by Boersma and Weenink on 1995. It can run on a wide range of operating systems including, Linux, UNIX, Mac and Microsoft Windows. Its latest version is PRAAT 5.3.39. Praat extracts as many as 30 parameters by its voice analysis. However, we have considered F0, jitter, Shimmer and HNR only since many other studies have proved their efficacy[12].Maryn et al[6] (2009:217) as quoted by Williamson[7] compared frequency perturbation (jitter) and amplitude perturbation (Shimmer) measures using both MDVP and PRAAT programs, and both a purpose-built recording system and a personal computer-based system for acoustic voice assessment. Theauthor noted that MDVP consistently yielded higher measures than PRAAT and concluded that "one can hardly compare frequency perturbation outcomes across systems and programs and amplitude perturbation outcomes across systems..."PRAAT software itself can calculate five different measures of jitter and six different measures of Shimmer. However, Graham Williamson[7] et al in his articlestate: "It is difficult to be precise about norms for acoustic measures such as jitter, Shimmer, noise-to-harmonics ratio and fundamental frequency. There are many factors which militate against declaring all-encompassing norms. Some of these are person-specific (e.g. gender and age differences), cultural (e.g. what north Americans may consider to be within normal limits may be different from what north Koreans consider to be typical), and related to the testing environment (e.g. variation in the equipment used, and – importantly – the use of different algorithms in the software programs which are used to make the measurements). Measures of jitter and Shimmer using one software program cannot always be compared directly with measures made by another software program" The following table gives comparison of different values from studies: Bielamowicz et al[13] have compared perturbation measures from C Speech, Computerized Speech Laboratory, SoundScope, and a hand marking voice analysis system and state that "Measures of perturbation in the various analysis packages use different algorithms, provide results in different units, and often yield values for voices that violate the assumption of quasi-periodicity.AA Simone et al[8]state "We assume that comparison of results collected with different vocal acoustic analysis programs can present differences even when using similar measures, owing to differences in algorithms, methods to calculate fundamental frequency, type of microphone used, type of storage of the recorded voice and type of token used, if connected or sustained speech" Areis, et al[11] opine that despite the accuracy and reliability of each machine, authors have agreed to standardize normative data individually due to a number of factors that may cause variations among each system. These possibilities include the type of programming of the acoustic analysis software, the use of recording criteria, type of microphone and other devices used in voice recording. Despite these complications, some authorities do declare so-called thresholds of pathology. For example, the Multi-Dimensional Voice Program (MDVP)(Kay Elemetrics, 2008) indicates a threshold of pathology of <=1.040% for jitter and <=3.810% for Shimmer.The normative Indian voices data was also given by Hemaet Al[14]using MDVP software. About 104 voice samples were used with 54 males and 43 females. The parameters were compared with western voices which showed differences in perturbation measurements. They concluded that these differences are due to change in vocal tract length, mass and tension of vocal tracts.There is a growing international trend for significant technological developments in the field of voice and speech evaluation, especially in the advancement of vocal acoustic analysis software. For this reason, standardization of normal acoustic measures is necessary due to the variation of systems protocols and software algorithms[5].Several acoustic analysis softwares have demonstrated normal and pathological voice conditions. Despite the accuracy and reliability of each machine, authors have agreed to standardize normative data individually due to a number of factors that may cause variations among each system. ANA et al[5] felt that standardization of fundamental frequency measures (f0), jitter, Shimmer and harmonic noise ratio (HNR) for young adults with normal voice is the need of the day. They studied normal voices of 20 males and 20 females. between 20 and 45 years, without signs and symptoms of vocal problems using CSL-4300 Kay-Elemetrics software with vowels /a/ and /é/[2]. Their Results showed that for females, vowels /a/ had average measures of f0 as

205.82 Hz; jitter of 0.62%;Shimmer of 0.22 dB ; HNR of 10.9 dB , respectively. For males, vowel /a/ had average measures of: f0 119.84 Hz; jitter of 0.49% ; Shimmer of 0.22 dB ; HNR 9.56 dB respectively. Both f0 and NHR female measures were significantly higher than their male counterparts. They felt significant differences with other studies. Their findings are given as mean only without range.Acoustic measures for normal adult voices are given by Williamson[7] from PRAAT Software as follows: Percentage of Jitter and Shimmer is to be less than 1.04% and 3.8% respectively. HNR Value is to be less than 20 Db. Fo in males and females as less than 128 and 225 Hz respectively. These normative values are similar in our study. The value of Jitter in our study was 0.018% (Jitter ddp) which is very low when compared to other studies like Bonzi et al[12] of local jitter percentage of 0.36%. Wang[10] was 0.56%, Simone et al[8] was 0.37% (jitter ratio), Kirt et al[11] was 0.46%, Fillippe et al[5] was 0.62% (jitter Avg). When the value of Shimmer (0.1572) was compared with other studies like A Simone et al[8] (2.37 db), Bonzi et al[12] (local Shimmer%, 3+-0.9) and Toranet al[9] (1.6%), it is found to be low. But studies like Horii[16] (0.132db), Ana et al[5] (0.22db) Aries et al[11] (0.23%) showed similar values.On analysis of HNR value from our study (20 db), it was similar to Bonziet al[12] (20+/-2 db) and Fernandez Leisa et al[15] (18db). It was higher value of 25.81 db in KC Toran et al[9]and very less in ANCFillippeet Al[5] (9.56 db).The comparision of values of parameters show significant differences which necessitate standardization.The purpose of this paper was to state the significance of voice analysis systems in the prediagnosis of certain medical conditions which later may transform into fatal or incurable diseases. The proposal of a voice database, standard voice analysis tool, and method of voice measurement is done. This methodology ensures accuracy, patients' ease, economy and less time consuming in predicting symptoms at early stages. It can be used by any medical practitioner without prior training. Since it is not possible to have a single world standard for normal voice parameters every clinic or laboratory should develop its own standards and use these as benchmark for further comparisons or clinical evaluations.

CONCLUSION

Acoustic analysis of voices is a simple noninvasive technique which can be effectively carried out easily at any clinic with good accuracy and reasonable prediction of symptoms. It is one of the objective tool for characterization of normal voice that can be used by a clinician. Jitter, Shimmer and Harmonic noise ratio are the parameters measured whose ranges give an idea of normality of the voice.

CONFLICT OF INTEREST

Conflict of interest declared none.

REFERENCES

1. Berg VDJ. Myoelastic-aerodynamic theory of voice production. J of Speech and Hearing Res. 1958; 3(1): 227-244.
2. Herbert DH, Dunker E. Husson's Theory; An Experimental Analysis of his Research Data and Conclusions. Arch Otolaryngol, 1967;85(3):303-313.
3. Praat:Doing phonetics by computer.Paul Boersma and Weeninck. Available from http://www.fon.hum.uva.nl/praat/
4. Parsa V, Jamieson DG. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. J Speech Lang Hear Re., 2001;44(2):327-39.
5. de Felippe AC, Grillo MH, Grechi TH. Standardization of acoustic measures for normal voice patterns. Brazilian journal of otorhinolaryngology. 2006 Oct 31;72(5):659-64.
6. Maryn.Y. Acoustic measurement of overall quality of voice of sustained vowels and continuous speech. 2010[Internet] Available from https://biblio.ugent.be/publication/888156/file/888179.pdf
7. Grahaam Williamson. Acoustic Measures (Norms). Voice. [internet] [cited 01 February 2016]. Available from http://www.sltinfo.com/acoustic-measures-norms/
8. Simone AA. Marcos G. José CP, Marcelo OR.Brazilian Journal of Otorhinolaryngology,2002. 68:4(14), 540-544 availablefromhttp://oldfiles.bjorl.org/conteudo/acervo/acervo_english.asp?id=446
9. Toran KC, Lal BK. Objective analysis of voice in normal young adults. Kathmandu University Medical Journal, 2009; 7(4): 374-377.
10. Wang CC et al. Voice Acoustic Analysis of Normal Taiwanese Adults. J Chinese Med Assoc. 2004; 67(4):179-184.
11. Kirt AD, Ray C, Fredrick YH. Vocal acoustic measures of Asymptomatic Filipino young adults at a private tertiary Hospital in Quezon city- A pilot study. Philippine J of Otol and H and N Surg. 2012 JulyDec;27(2):12-17.
12. Bonzi EV, Grad GB, Maggi AM. Study of the characteristic parameters of the normal voices of Argentinian speakers.papers in Physics.2014;6 (1)23-25.
13. Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS. Comparison of voice analysis systems for perturbation measurement. J of Spch and Hear Res. 1996 Feb;39(1):126-34.
14. Hema N.; Mahesh, Sangeetha; Pushpavathi, M. Normative Data for Multi-Dimensional Voice Program(MDVP) for Adults - A Computerized Voice Analysis System. J of the All India Institute of Spch& Hear. 2009; 2 (28): 1-5.
15. Fernández LR et Al, Acoustic analysis of the normal voice in nonsmoking adults. Actaotolaryngolespanol. 1999;50(2):134-4.

Reviewers of this article

Dr.R.M.Karadi, M.S. (ENT)

Prefessor & HOD, Department of ENT,
Shri B. M. Patil Medical College, Bijapur,
India

Asst.Prof.Dr. Sujata Bhattacharya

Assistant Professor, School of Biological and Environmental Sciences, Shoolini University, Solan (HP)-173212, India

Prof.Dr.K.Suriaprabha

Asst. Editor , International Journal of Pharma and Bio sciences.

Prof.P.Muthuprasanna

Managing Editor , International Journal of Pharma and Bio sciences.

We sincerely thank the above reviewers for peer reviewing the manuscript