

# On the Differences in Prosodic Features of Assamese Speech Based on Types of Emotions

Somi Kolita, Purnendu Bikash Acharjee



**Abstract:** *Speech is a vocal communication through which we achieve the information about speaker, language, message etc. Speaking is the process of converting discrete phonemes into continuous acoustic signal. Speech carries linguistic information that is associated with emotions, along with vocal information, which can be extracted by speech processing methods. In this paper, we basically analyse on prosodic features of different emotional speech in Assamese language for identify the comparison between male and female speakers who linguistically react differently in the same situations. We also apply statistical paired t-Test on prosodic features for showing the significant difference between male and female speech in different emotions.*

**Keywords:** *Vocal Communication, Assamese language, Linguistically React Differently.*

## I. INTRODUCTION

Speech is the most important bio signal that human being can produce and perceive. The prosody plays a particular role in emotional speech. The prosodic features adding sense in speech providing emotions of speaker.

In addition, every individual speaker has their own style of speaking. The characteristics like articulation rate, intonation habit and loudness differ in case of different speakers. The factors that affects one's emotion that is expressed and inferred in a speech are gender, health, age, social status, education, physical engagements etc.

In previous researches, most have engaged in recognizing emotion in speech automatically and also have tried to incorporate this technology into real world applications. Because, identifying the features which describe the emotional content in speech is a difficult task.

Emotional speech analysis is carried out at various levels like segmental, sub segmental and supra segmental. The performance analysis of emotion recognition is mainly depends on the speaker and the phonetic information. Emotional features are broadly classified as spectral features and prosodic features.

Most emotional speech can be attributed to the larger segments known as supra segmental features or prosodic features.

The prosodic features are rhythmic, intonational and speaking rate properties in speech. Prosodic features mainly estimated over the uttered speech sentence in the form of long term statistics [1].

Prosody features are resultants of vocal fold activity and the spectrum peaks which are influenced by the vocal tract. The prosodic features gives the utterance level characteristics.

Prosody is rich source for understanding about speech. It has great significance in emotional speech analysis [1, 3]. In recent years, many researcher has done more computational work towards prosodic modelling for automatic speech

Recognition system and understanding. The basic idea of the paper is to focus on the study of the basic correlation between

Prosodic features like Pitch, intensity and formant emotional characteristics. Based on these prosodic characteristics we revealed the significant difference between male and female speech sample using statistical paired t-Test. According to the paired sample t-test, statistical significance difference between two sample mean occurs if the p value (probability of observing) is less than equal to  $\alpha = 0.05$ .

## II. METHODOLOGY

### 2.1. Data Collection

In this current analysis, the vowel, consonants and words sounds of the Assamese language for both female and male have been collected for analysis.

For this purpose, some male and female speakers have been picked from a college of Jorhat district in age group of 20-23.

Samples consisted of simulated speech of one day to day emotionally biased-sentence in Assamese language, which repeated three times with three emotional styles-Angry, Sad and Surprise.

The recorded Assamese sentence for analysis is *Moi Ghorot jam* (মই ঘৰত □□□ ||). Then, I divided the sentence consisting of three words into 3 different parts using Audacity i.e. -মই – First word, ঘৰত – Second word and □□□ – Third word.

Following is the figure1 and figure 2 represents the waveforms for female and male speech sample of the sentence (মই ঘৰত □□□ ||) in Assamese language with different emotion types (a)Angry (b)Sad and (c)Surprise.

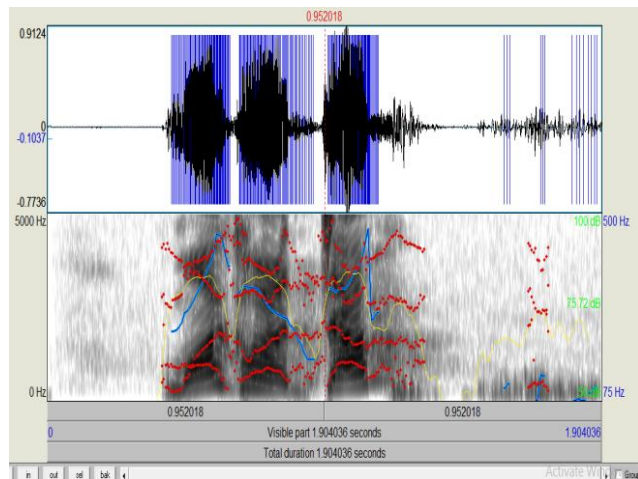
Revised Manuscript Received on February 28, 2020.

\* Correspondence Author

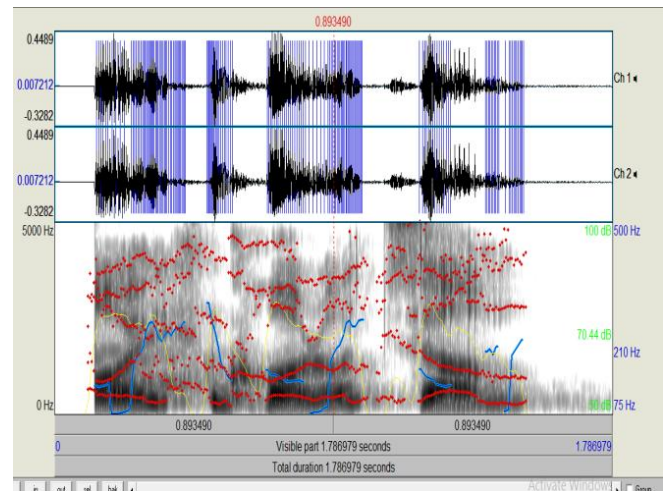
**Mrs. Somi Kolita\***, Assistant Professor, Department of Computer Science, Bahona College, Jorhat, Assam, India.

**Dr. Purnendu Bikash Acharjee**, Assistant Professor and Associate Dean, School of Computing Sciences, The Assam Kaziranga University, Assam, India.

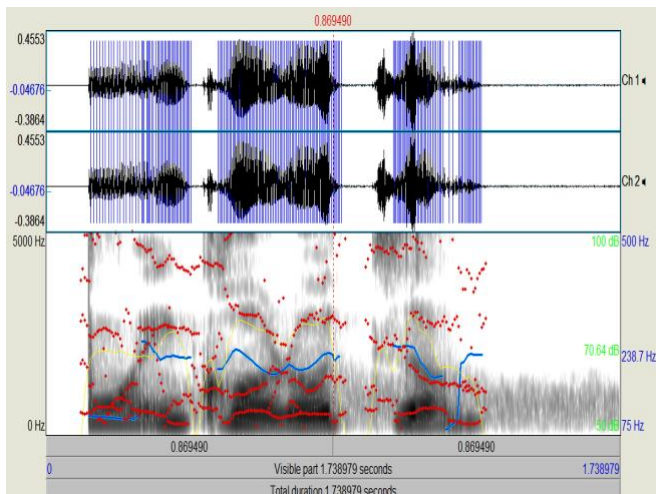
© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>



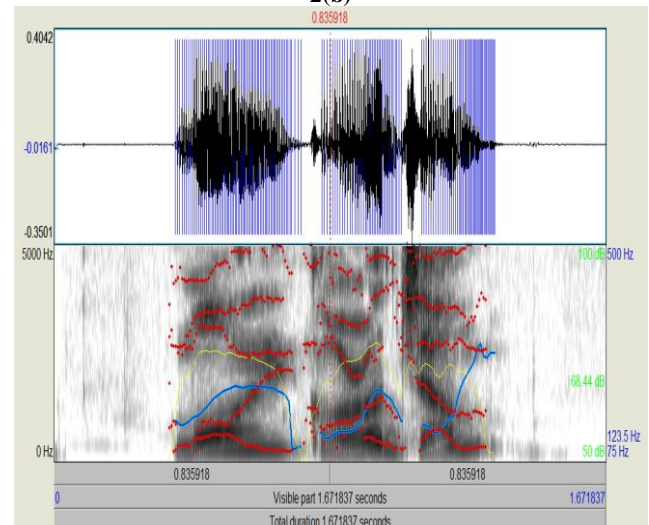
1(a)



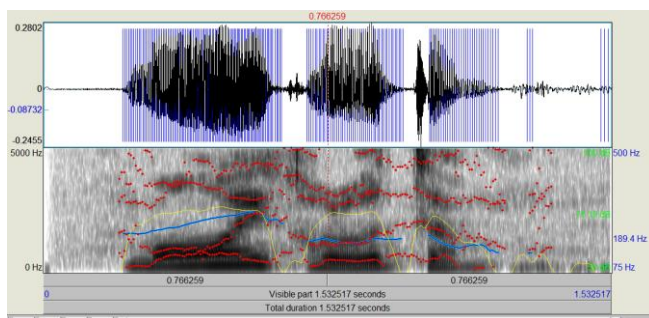
2(b)



1(b)

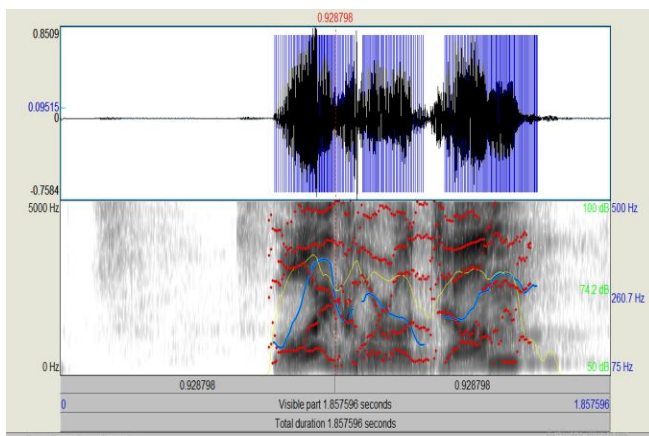


2(c)



1(c)

Figure1: Female Speech sample of Assamese sentence **Moi Ghorot jam**( মই ঘৰত ) with three different emotions.



2(a)

Figure2: Male Speech sample of Assamese sentence **Moi Ghorot jam**( মই ঘৰত ) with three different emotions.

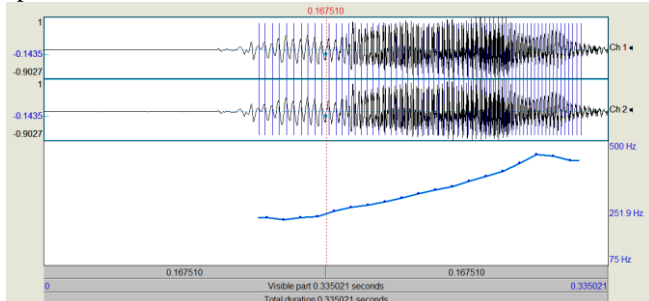
## 2.2. Prosodic parameters

Since most of research on emotion analysis/recognition is based on features and classification based approaches. Feature selection determines the features which are most beneficial for analysis /recognition because most of the feature classifiers are negatively influenced by irrelevant features. Feature based approaches concentrates on analysing speech signals and effectively estimating feature parameters for representing human emotional states. But the classification based approaches mainly focuses on designing a classifier to determine distinctive boundaries between different emotional states of human speech.[2]

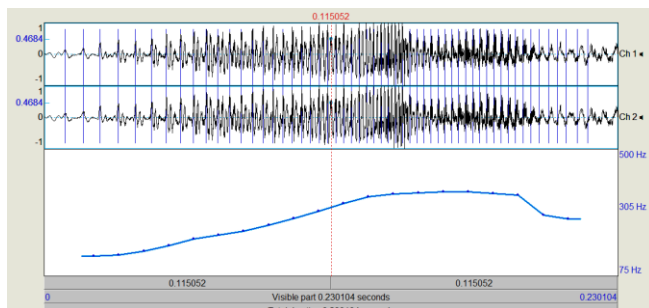
The prosodic features like pitch, intensity (loudness), duration, speaking rate and voice quality are more important to identify different types of emotions. In our work, We have consider three prosodic features like intensity, pitch and formant for our analysis and extracting these features from collected Assamese speech using praat speech analysis software. Pitch: Pitch is an important acoustical parameter of speech analysis. It represents the vibration of Vocal cord during the sound production. Based on the vibration, sounds are classified voiced and unvoiced.



Basically pitch is also known as the fundamental frequency of sound which are provided with glottal pulse. Sound can be characterized based on pitch value, intensity or loudness like comparison of low or high in musical sounds. Generally, the pitch range for male voices is 60–180 Hz, and the pitch range of female voices is 160–300 Hz, but this range changed depending on the types of emotion and language. Following figure 3 shows the pitch counter of Assamese word(মই) from the dataset in angry emotion of both (a)male and (b)female speakers

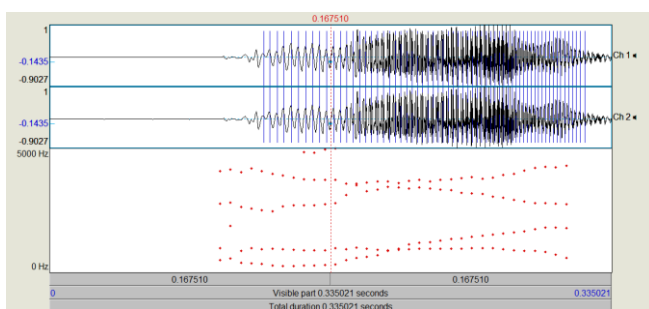


3(a)

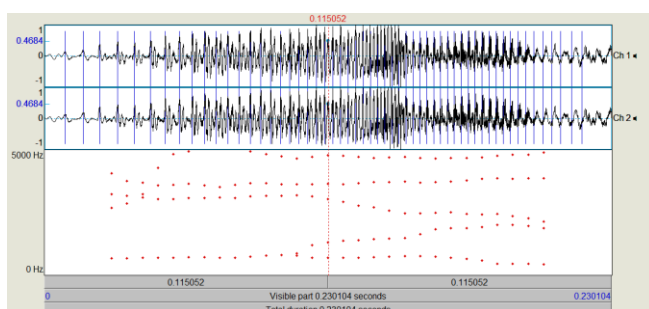


3(b)

Formant: Formant frequencies represent the peak of frequency response of the vocal tract filter. Formant are the components of acoustic signal that produced by sound. Following figure 4 shows the formant of Assamese word(মই) from the dataset in angry emotion of both (a)male and (b)female speakers.

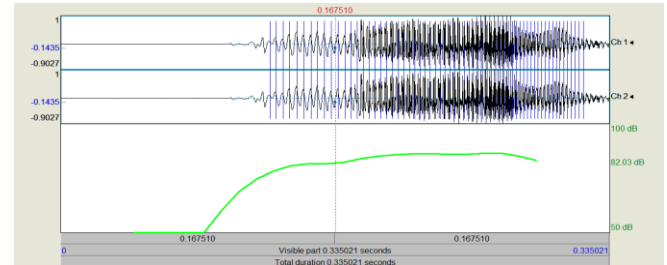


4(a)

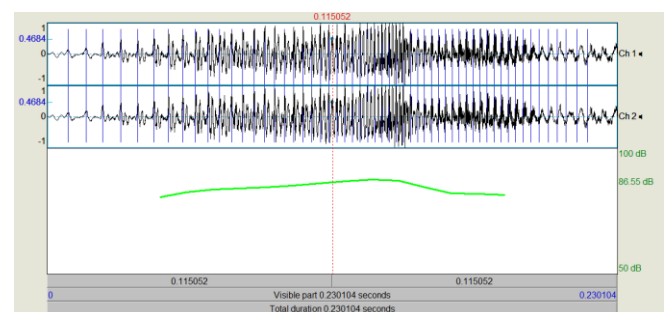


4(b)

Intensity: Intensity or vocal intensity of the speech signal refers to the loudness effect of speech signal. Vocal intensity is related to the sub glottis pressure of the airflow, which depends on the tension and the vibrations of the vocal folds. Following figure 5 shows the intensity of first Assamese word (মই) from the dataset in angry emotion of both (a)male and (b)female speakers.



5(a)



5(b)

### 2.3. Statistical Procedure

The data were subject to the statistical procedures of Two paired t-Test, in which the two groups were compared. These statistics were used so as to indicate the significant difference between the prosodic features of male and female speaker' s in same situation with different emotions.

## III. RESULTS AND DISCUSSION

We have collected 270 speech samples, 90 for each of the three assamese words (মই | ঘৰত | □□□). Each word is spoken by each speaker in three types of emotions: “angry”, “sad” and “surprise”. Extracting the prosodic features from these collected samples using praat which is a speech analysis software tool. Thus, we have built the Database which contains all the speech samples and along with their corresponding prosodic features. We enumerated and distinguished the prosodic means from all the samples and the results are given below:

Following figure 6 represents the graphical representation of comparison between mean pitch for male and female speakers and figure 7 gives the comparative graph of mean intensity of speech signal for male and female speakers .Figure8 provides the graphical representation of Mean Formant Frequency comparison between Male and Female. Table 1 and Table 2 gives the average mean values of the prosodic features which says that Female pitch is highly increased than Male in three emotions. Similarly, in Angry emotion formant frequency is increased than male but in Sad and Surprise,

## On the Differences in Prosodic Features of Assamese Speech Based on Types of Emotions

formant frequency of male is nearly increased than female speaker. But, the intensity of male in Surprise are slightly greater than female. We also apply Paired t-Test on the mean average of three prosodic features of male and female speech samples. Based on the Paired t-Test results, it was shown that the two male and female sample were totally different from one another. Then we give significant difference between prosodic mean of male and female speech, because in each three paired test  $p < .05$ .

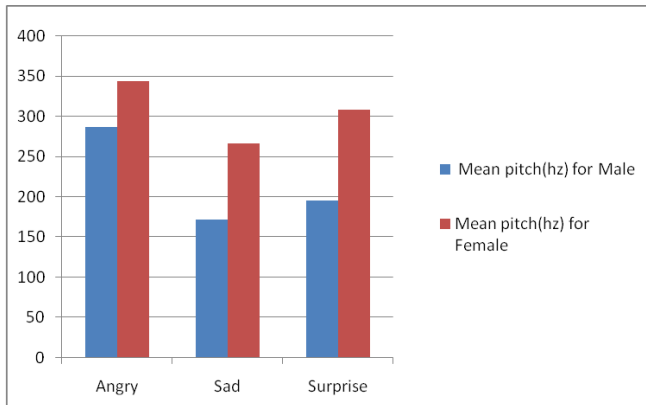


Figure 6: Comparison of average Pitch of Male and Female speakers

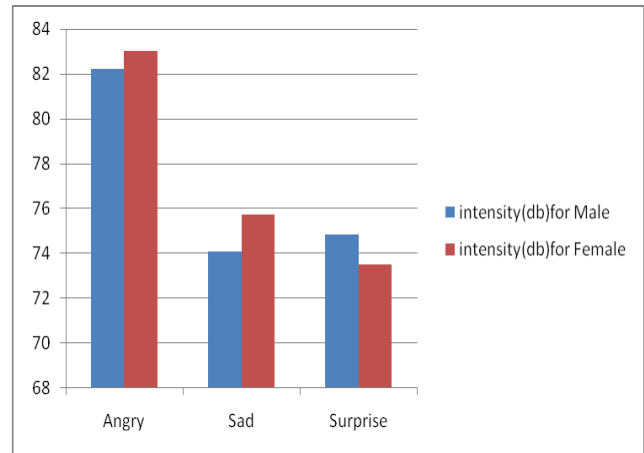


Figure 7: Average intensity comparison of Male and Female speakers

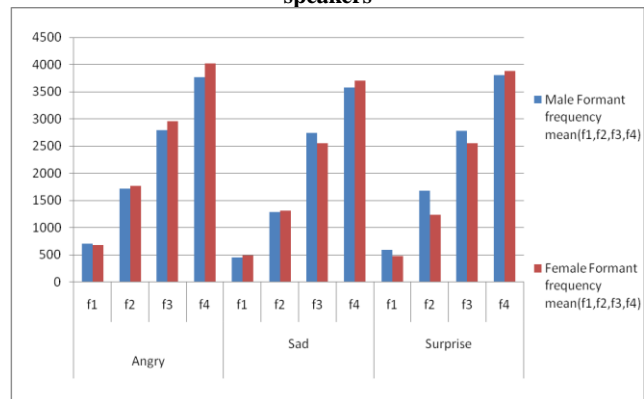


Figure 8: Average Formant Comparison of Male and Female speakers

Table 1: For Male

	Time(sec)	Mean formant frequency(hertz)				Pitch mean	Pitch variance	Intensity mean
		f1	f2	f3	f4			
Angry	0.170478	711.760	1720.209	2799.643	3767.382	287.036	3187.015	82.229
Sad	0.192669	448.765	1294.595	2740.515	3573.001	172.508	2477.731	74.107
Surprise	0.16076	591.342	1685.632	2786.611	3800.114	196.1711	4883.651	74.853

Table 2: For Female

	Time(sec)	Mean formant frequency(hertz)				Pitch mean	Pitch variance	Intensity mean
		f1	f2	f3	f4			
Angry	0.169712	677.562	1771.162	2954.561	4022.171	344.223	9657.42	83.020
Sad	0.185649	497.783	1317.625	2551.163	3697.442	267.089	5347.82	75.744
Surprise	0.18484	485.064	1236.391	2551.731	3877.507	308.179	6276.14	73.503

## IV. CONCLUSION

Summary of the results from our analysis as follows: 1. For angry speech which uttered by the female speakers highly increases compared to that of normal speech. 2.

Pitch and formant of female speaker is highly increases than male in angry emotion and also observed the significant difference between male and female speech.3. Intensity of female speaker slightly decreases than male speaker in surprise emotion.

Results of implementation shows the possibility in which our approach is effective for enhancement in Human Computer Interactions. We will be try to implement the future works are to collect more speech samples with emotion labels on them and try to recognize the emotion using classifiers with reference to Assamese language. It is necessary for us to reconsider the process of estimating emotion in speech based on prosodic experiments. For example, the features on which we will focus and the combination of features and classifiers that we will use. Then we have to evaluate our approach. Naturally, human beings express multiple numbers of emotions in their speech. So in order to improve human computer interaction we have to think on processing multi emotions in speech.

twenty five research publications in international journals and refereed conference proceedings.

## REFERENCES

1. X.Arputha Rathina et al, "Basic Analysis on Prosodic Features in Emotional Speech", International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.2, No.4, August 2012
2. Shiva Prasad K M.et al"Speaker Dependent Speech Feature Based Performance Evaluation of Emotional Speech for Indian Native Languages"International Journal of Computer Science and Information Security (IJSIS), Vol. 14 CIC 2016
3. Laya Heidari Darani, Hourieh Heidari Darani et al "Language and gender: A prosodic study of Iranians talks", Procedia - Social and Behavioral Sciences 70 ( 2013 ) 423– 429
4. Jordi Adell, Antonio Bonafonte et al, "Analysis of prosodic features: to-wards modeling of emotional and pragmatic attributes of speech", Proc. Natural lang proc, 2005.
5. Takeda.S, Ohyama G, and Tochitani.A, "Diversity of Prosody and its Quan-tative Description" and example: analysis of "anger" expression in Japanese Speech, , Taejon, Korea,pp.423-428,2001.
6. Jordi Adell, Antonio Bonafonte et al "Analysis of prosodic features: towards modelling of emotional and pragmatic attributes of speech, Procesamiento del Lenguaje Natural, núm. 35 (2005), pp. 277-283
7. Yasuki Hashizawa et al"On the Differences in Prosodic Features of Emotional Expressions in Japanese Speech according to the Degree of the Emotion", Speech Prosody 2004 Nara, Japan March 23-26, 2004
8. Sallar Khan et al, "Analysis of Children's Prosodic Features Using Emotion Based Utterances in Urdu Language", Engineering, Technology & Applied Science Research Vol. 8, No. 3, 2018, 2954-2957
9. Tarunam Chaudhary et al, "Pitch, Formant Frequencies, Intensities as Speech Cue to perceived age," International Journal Of Engineering And Computer Science , ISSN:2319-7242 Volume-4 Issue 7, July 2015
10. Jordi Adell, Antonio Bonafonte et al, "Analysis of prosodic features: towards modeling of emotional and pragmatic attributes of speech", Proc. Natural lang proc,2005.

## AUTHORS PROFILE



**Mrs. Soma Kolita**, pursuing Phd from School of computing Sciences, The Assam Kaziranga University, Jorhat on speech emotion recognition area and completed Master of Computer Application from Dibrugarh University and currently working as a Assistant Professor in Dept of Computer Science, Bahona College, Jorhat, Assam. Has four research publication in national and international journals.



**Dr. Purnendu Bikash Acharjee**, Assistant Professor and Associate Dean in School of computing Sciences, The Assam Kaziranga University. M.Sc in Computer Science, M.Tech in Information Technology, PhD under Faculty of Technology from Guwahati university. Senior Research Coordinator, sponsored by MCIT, Govt. of India in 2012. Has four publications in the IEEE Conferences and more than