UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002.

Annexure -VII

PROFORMA FOR SUBMISSION OF INFORMATION AT THE TIME OF SENDING THE FINAL REPORT OF THE WORK DONE ON THE PROJECT

1. NAME AND ADDRESS OF THE PRINCIPAL INVESTIGATOR : **Rizwan Rehman , Assistant Professor, Centre For Computer Studies, Dibrugarh University, Dibrugarh.**

2. NAME AND ADDRESS OF THE INSTITUTION: **Dibrugarh University, Dibrugarh, Assam-786004**

3. UGC APPROVAL NO. AND DATE: **F.No. 41-1362/2012(SR) , 30-July-2012**

4. DATE OF IMPLEMENTATION : **09/11/2012**

5. TENURE OF THE PROJECT : **Two Years**

6. TOTAL GRANT ALLOCATED : **1,95,000**

7. TOTAL GRANT RECEIVED :   **1,45,000**

8. FINAL EXPENDITURE : **1,41,500**

9. TITLE OF THE PROJECT :   **"An Improved Mishing grapheme to phoneme conversion using statistic and linguistic information"**

10. **OBJECTIVES OF THE PROJECT** :
    1. Explore the structure of *Mishing* (Shaiyang Miri) Language.

    2. Creating a repository of *Mishing* Words.

    3. Analyse the vowels in *Mishing* Language using Formant Analysis, Spectrogram, 19-channel auditory filter bank and through normalization of vowels.

    4. Creating a corpus of vowels and consonants for grapheme to phoneme conversion.

11. WHETHER OBJECTIVES WERE ACHIEVED :**YES**
(GIVE DETAILS): : **ENCLOSED**

12. ACHIEVEMENTS FROM THE PROJECT: **ENCLOSED**

13. SUMMARY OF THE FINDINGS : **ENCLOSED**
(IN 500 WORDS)

14. CONTRIBUTION TO THE SOCIETY
(GIVE DETAILS): : **ENCLOSED**

15. WHETHER ANY PH.D. ENROLLED/PRODUCED : **NO**
OUT OF THE PROJECT

16. NO. OF PUBLICATIONS OUT OF THE PROJECT: **ONE**
(PLEASE ATTACH RE-PRINTS): **ATTACHED**

(PRINCIPAL INVESTIGATOR)

(REGISTRAR/PRINCIPAL)

Registrar
Dibrugarh University

ACHIEVEMENTS FROM THE PROJECT

1. Successful recognition of vowels in SHAI`YÂNG MIRI (*Mishing*) language.
2. Repository of SHAI`YÂNG MIRI (*Mishing*) Words
3. Creation of repository of vowles , consonents and words of SHAI`YÂNG MIRI (*Mishing*) language.
4. Efficient conversion from grapheme to phoneme.

SUMMARY OF THE FINDINGS :

SHAI`YÂNG MIRI also known as MISHING language is spoken by the Mishing people residing mainly on the banks of Brahmaputra river. The origin of SHAI`YÂNG MIRI language is Tibeto-Burman language spoken by more than 500,000 people residing in Lakhimpur , Dhemaji, Sivasagar, Jorhat, Golaghat and Tinsukia district of Assam. SHAI`YÂNG MIRI language, in absence of its own script uses the Roman script for its lexicographical determinants. Therefore, there is a difference between the spoken form and the written form.

Based on the type of sound produced it is found that SHAI`YÂNG MIRI (Mishing) language consists of 41 phonemes of which 16 are consonants and 15 vowels including 10 allophones [6]. The vowels include /a/, /e/, /i/, /o/, /ô/, /û/, /ee/, /ea/. These eight vowels are the cardinal vowels with their seven long variances which are /ah/, /eh/, /ih/, /oh/, /uh/, /eyh/, /iuh/.

Table 1. Cardinal Vowels

| Phoneme	Place of Articulation	Example
/a/	Front open	Esar
/e/	Central half open	Ta-peta
/i/	Front close	Moi`ya
/o/	Half open	Ko`pak
/ô/	Back half close	Tat-p ôa
/u/	Back close	Bur
/ea/	Half close	Geelean
/ee/	Central close	Leegang

The study was performed by recording the 8 cardinal vowels in Shai`Yâng Miri (Mishing) language spoken by 10 male and 10 female native speakers. The recording was done using Speech Filling System (SFS) package and then vowels were analyzed by measuring the formant of each vowel by using fmanal application of SFS package. This analysis provides a proper match between the tracks and the spectrogram. Formant values thus obtained are converted to BARK scale.

Table 2. Mean formant frequencies of phonetic vowel as produced by 10 male and 10 female native speakers

VOWELS	FORMANTS	
	F1 / Hz	F2 / Hz
/i/	368	2315
/e/	403	2119
/ea/	460	1720
/ee/	581	1551
/a/	751	1255
/o/	558	1102
/ô/	380	1080
/u/	316	1065

Table 3. Normalized Vowels of Shai`Yâng Miri (Mishing) language

VOWELS	FORMANTS	
	F1 (Bark)	F2 (Bark)
/i/	3.708	13.988
/e/	4.042	13.398
/ea/	4.566	12.000
/ee/	5.600	11.313
/ a /	6.897	9.935
/o/	5.411	9.119
/ ô /	3.824	8.995
/u/	3.192	8.909
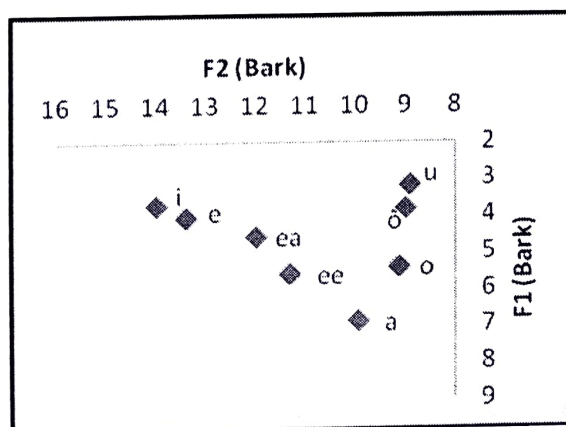
"vowel quadrilateral", with formant frequencies converted into Bark Scale

The investigation is completely based on formant analysis, spectrogram analysis, analysis of 19-channel auditory filter bank output and normalization of vowels with respect to the formants F1 and F2 i.e. the first and the second formant. The LPC method is used for determining the frequencies and amplitudes of formants in speech.

The normalization points for each of the 8 cardinal vowels in Shai`Yâng Miri (Mishing) language have been found. The results were examined with the help of the SRS software for the recognition of vowels in continuous speech with a success rate of around 88-95%.

Table 4. Recognition of Vowels In All Sequence

Vowel	Number of vowels in text	Recognized	Not Recognized	Percentage
/a/	40	35	5	87.5
/e/	41	37	4	90.24
/i/	39	35	4	89.74
/o/	37	33	4	89.19
/ ô /	35	32	3	91.43
/u/	36	33	3	91.67
/ea/	20	17	3	85
/ee/	18	16	2	88.89

CONTRIBUTION TO THE SOCIETY

This will help the people of *Mishing* society in two fold manner i.e. primarily it will enable to create systems that can recognize the *Mishing* language hence enabling them to interact with the computer systems, say in reading a text, this can be applied in many different applications also . Secondly, it will allow their literature to be computerized and recognizable by various search engines which is now difficult at present as all literatures are in image format. Further, this technique can be applied to various other languages of the north-eastern region.