International Journal of Computer and Communication Technology

Volume 3 | Issue 1 Article 10

January 2012

Markov Model Based Oriya Isolated Speech Recognizer-An Emerging Solution for Visually Impaired Students in School and Public Examination

Sanghamitra Mohanty
Dept of Computer Sc. & Application Utkal University Bhubaneswar, Orissa, India,
sangham1@rediffmail.com

Basanta Kumar Swain
Dept of Computer Sc. & Application Utkal University Bhubaneswar, Orissa, India, technobks@yahoo.com

Markov Model Based Oriya Isolated Speech Recognizer-An Emerging Solution for Visually Impaired Students in School and Public Examination

Sanghamitra Mohanty[1], Basanta Kumar Swain[2]
Dept of Computer Sc. & Application, Utkal University, Bhubaneswar-751004, Orissa, India
sangham1@rediffmail.com[1], technobks@yahoo.com[2]

Abstract--**In this paper we have tried to focus on problems faced by visually impaired students of Orissa in school and public examinations and discussed the emerging solutions to those problems at the end in order to minimize those problems. Assessment plays an important role in the education system for evaluating student's skill in that area. Assessment can be carried out by allowing students to appear exam in terms objective or non-objective type questions but due to the advantages offered by objective type assessment like offer speedier marking, reduced need for cross marking, greater fairness (reliability), more coverage of the syllabus and finally automate the marking process which enforces examination system to be held as objective rather non-objective type. Therefore, it is mostly implemented in examination system conducted in public or school exams. But it is very regret to mention that this assessment procedure is not well suitable for students those are visually impaired as they are unable to read what they have written or it is not possible to know if the ribbon has run out or if the line spacing is faulty unless the supervisor tells them. These are the main obstacles faced by visually challenged people which lead to disqualification for higher studies that ultimately reason to failure of motto i.e. "Education for All". In this research study we have considered close ended objective type question. All possible responses of the closed questions are designed such a way that they are mutually exclusive. The closed ended questions are of five types namely fill-in-the-blanks questions, dichotomous questions, ranking scale questions, multiple choice questions and rating scale questions. As all the answers to the above types of questions carry isolated answers, we have developed Oriya isolated speech recognition system which can easily convert isolated answers uttered by visually impaired students to isolated text(isolated speech to isolated text). Hence, visually impaired students can answer the question as normal student because they do not need at all their hand and eyes to respond the answer. As a result of which overcomes the problem faced by visually challenged people to answer the questions without help of keyboard. Oriya isolated recognition system is developed using hidden Markov Model by incorporating1800 isolated Oriya words collected from 30 speakers in training stage. Testing stage is carried out by considering 5 speakers which yields 76.23% word accuracy on seen data and 58.86% on unseen data.**

1. INTRODUCTION

India is a country physically divided into 30 States and 7 Union Territories hosting immense diversity of traditions, languages, castes and cultures. Orissa is one of the 30 states in India, situated in the eastern part along the coast of Bay of Bengal. The total population of the State is 3.680 crores as per 2001 census. The multilingualism and multicultural society of India also presents many models of education especially from the perspective of education of learners with disabilities. Education has recently become a fundamental right for children between 6 – 14 years of age (86th Amendment Bill,2002). Broadly speaking, India has invested a lot of effort and money in creating a vast network of provisions for education of every level. However these efforts, in spite of being targeted to all children have failed to bring the disabled population fully under the fold of education. It is hard to believe that there are 45 million blind people in the world, 12 million live in India, and 180 million visually impaired people in the world, 52 million live in India!. The prevalence of blindness in India is estimated to be 1.5 per cent of the population. According to Census 2001, total population of Orissa is 36,804,660, out of this 1021335 persons are living with disability and 5.13 lakhs are visually impaired. Visual impairment becomes plague to educate the visual impaired students in spite of good attempts by state government, non-government or voluntary organization to provide basic education to those students. It is being well visualized that development in education along with other allied sectors would play a vital role to bring about desirable changes in the country. That means visually challenged people of India must be provided opportunity in education to develop critical thinking and self determination and contribute to the progress of the country. The literacy percentage of the state Orissa is 63.61% in which male and female literacy stands as 75.95% and 50.97% respectively (Source census 2001, Provisional)[1]. A lot of attempts have been made to mainstream the visual impaired children in education process. Most importantly by providing free textbooks, opening of new schools, hostels, teachers through mass teachers training on child centered approach. However, there seem to be lack of achievement in retention and achievement of students (on the basis of completion rates) in schools. The root cause for exclusion of students(visually impaired) from education is due to the assessment procedures in schools which become uphill task to answer the question in spite of good knowledge in the domain of examination which forced visually impaired student from pursuing higher education. The same impedance scenario is also seen in public

examination for blind students to avail a job. Therefore, physically challenged students, if they are to have the adequate education promised them then they have to be assessed in ways which neither advantage nor disadvantage them vis-à-vis their non-handicapped peers. With this in mind, the particular problems that visually handicapped candidates currently encounter in examination are studied very carefully such as visually handicapped students need the exam paper in a form that they can read for themselves. This means braille for the braille readers, large or giant print for those who need it for partially sighted students. However, even where an Examination Board aims to provide brailed question papers, they may not actually reach all blind candidates. Even where braille papers arrive in time, mistakes in the braille or poor reproduction may cause problems for the blind candidates. Where no braille copy of the question paper is provided, the question paper must be read to the candidates. Furthermore, readers may be bored or impatient if asked to read something over and over again. Also, a reader's poor pronunciation or enunciation may mislead the candidate. In some situations the braille user wastes a lot of time searching for the option and the partially sighted and low vision candidates are having difficulties in scanning, are also very significantly slower and less efficient in looking from place to place on the page or in worse on different pages. These problems can be eradicated due to available of highly reliable text-to-speech(TTS) of Oriya language and other Indian languages[2]. Though text-to-speech became helpful to solve one side of the problems faced by the visual impaired students but other side problems remain unsolved that once candidate became fully clear about the questions with help of TTS then s/he wants to answer which should be captured in well efficient manner. Therefore, our ongoing research is towards development of digital solution based on hands free and eyes free mode so that visually impaired students and physically challenged students will not feel any kind of weakness in them as compared to their physically fit peers. We have developed isolated Oriya speech recognition system based on HMM which will be helpful to capture the answers judged by visually impaired candidates in public or school examination based on closed ended questions. On other hand we can say that answer is captured by converting isolated speech uttered by students into text which indicates the answer to respective question without typing explicitly by the students. In this research study the main objective is to provide fairness of assessment procedures for such visual impaired candidates to answer their examination papers without pen and key board as well as in more natural way so that it will to encourage students for higher studies which will reduce exclusion from schools as well as they can easily qualify public examinations in order to be employed. The implementation of Oriya isolated recognition system is based on hidden Markov model (HMMs) [3,4,5,6]. The training phase of the system is carried out by employing data collected from 30 speakers and another dataset of 5 people is used in test model for evaluating the recognition accuracy of the system. The vocabulary of the speakers was collected in the form of isolated word of Oriya language looking into the contexts of closed ended questions types such as fill in the blanks questions, dichotomous questions, ranking scale questions, multiple questions and rating scale questions. The system recognizes spoken answers in Oriya in the context of a assessment of students based on closed ended questions task. Thirteen dimensional MFCC values is being used as features in both training as well as testing stage.

2. ORIYA ISOLATED SPEECH DATABASE

Training and testing isolated Oriya speech recognition system needs isolated words which are collected from 30 Oriya speaker those are native to Orissa. The generation of a corpus for Oriya words and the collection of speech data are described below.

2.1 Word Corpus

Isolated words are collected by considering the closed ended objective type of questions of five categories. In our research study we have considered four questions of each type (category) which carries maximum four options. As result of which we have 60 isolated words (options) belonging to total 20 questions for five types of objective questions. In our study, we consider 30 speakers for training purpose which results 1800 isolated words for training the Oriya isolated recognition system. The details of Oriya isolated speech database is given in the Table 1.

Table 1: Details of Oriya isolated words

Type of questions	*Total numbers of questions of each type*	*Options of each questions*	*Total isolated words of each category*
Fill-in-the-blanks	*4*	*1*	*4*
Dichotomous	*4*	*2*	*8*
Ranking	*4*	*4*	*16*
Multiple Choice	*4*	*4*	*16*
Rating Scale	*4*	*4*	*16*

The following are the few samples of questions which is meant for collecting Oriya isolated words in our research work.

TYPE1: Fill-in-the-blanks question

Ex: i. /Ama desa……(kebe) swAdhinatA paithilA/(when did our country get freedom?)

ii. /Ama desara rAstrapatinkara nama…..(kana)/(who is our president of our country)

Type2: Dichotomous questions

Ex.: i. /khAibA purbaru hata dhoibA uchita ki/(should we wash hands before eating?/

a. Han(Yes) b.nA(No)

ii. /tamkhu sebana swastya prati hAnikaraka ki/(is tobacco injurious to health?)

a. Han(Yes) b.nA(No)

Type 3: Ranking type questions

Ex. i./tumara sikshakanka ra sikshAdAn upare tumara anubhuti kana/(what is your opinion about the teaching standards of teachers?)

a. /kharApa/(Poor) b. /chalaniya/(Fair)
c./bhala/(Good) d./atibhala/(Very Good)

Type 4: Multiple choice type questions

Ex: i. /Amara jAtiya pasu ra nAma kana/(what is the name of our national animal?)

a. /bAgha/(Tiger) b.mayura(peacock).
b. c./gAi /(cow) d. /ethiru kaunasiti nuhen/(none of these of above)

ii./ Amara rAstra bhAsa ….(kana)/ (what is our national language?)

a. /hindi/(Hindi) ./enrAge/(English)
C ./bangAli/ (Bengali) d. /odiA/(Oriya)

Type5: Rating type questions

Ex:/ehi mAsa re tume srenire kete thara anupastita rahichha/(how many time you are absent in the class this month?)

i./thare/(Once) ii. /dui thara/(Twice)
iii./bohu thara/(Many times) iv./Adau nuhen/(Never)

Oriya isolated words are collected using high quality, directional microphone at a distance of 5-10 cms from mouth in laboratory environment. The data collected from the desktop microphone contains noticeable noise. The speech data was sampled at 16 kHz and quantized with 16 bits.

3. FEATURE EXTRACTION

Feature extraction is a process to convert the speech signal into a sequence of vectors. The speech file was separated into frames by multiplication of overlapping Hamming windows. The interval was set to 10 milliseconds and the window length was 25milliseconds. In this research, the feature vectors consist of 13-dimensional Mel-Frequency Cepstral Coefficients are used in both training testing stage. Each of the feature vectors carries significant information about the spectrum and amount of energy in different frequency bands of a waveform at a given point in time. MFCC are calculated by passing the Oriya speech input signal through the sampling step, typical values are 16 kHz sampling rate with 16 Bit quantization. Next are the pre-emphasis and windowing processes where the Hamming window is used. The input signal then is divided into overlapped frames with the length of 25 milliseconds. Then, Fourier Transform of the signal is performed and mapping the log amplitudes of the spectrum derived onto the Mel scale using the triangular overlapping windows,. After that, DCT of the list of Mel log Amplitudes is taken which results MFCCs spectrum [6,7,8,9].

4. HIDDEN MARKOV MODEL (HMM) AS PATTERN RECOGNITION TECHNIQUE

HMM is one of the most powerful statistical pattern recognition techniques. In this research paper we used HMM for recognizing Oriya isolated word sequence uttered by human while answering the questions. Oriya isolated word recognition based on HMM consists of following components

- Feature extraction

Feature Extraction is to extract the features from the input signal or speech waveform which has been mentioned in Feature Extraction section.

- Acoustic Model

Acoustic Model provides statistical modeling for the acoustic observation sequence *O*. The model units can be based on semantically meaningful units, such as words, or phonetically meaningful sub-word units such as phonemes. In this research, the recognition unit is based on the phonemes.

- Language Model

Language Model provides linguistic and grammar constraints to the word sequence *W*. It is often based on the statistical N-grams language models. An N-gram is of the form $P(wn \mid w1, \ldots\ldots, wn\text{-}1)$, which is the probability

of observing word *wn* given the word history *w1, ……., wn-1*. However, language model would not be used in this system since this research is focused on the Oriya isolated word recognition. Therefore, this research will not focus on the linguistic and grammar limitations such as N-grams [10, 11].

- Dictionary

The dictionary is in the Sphinx dictionary format with distinct words entry. Dictionary contains the words available in language and the pronunciation of each in terms of the phonemes available in the acoustic model. As in Oriya language single word may be pronounced multiple ways. Multiple pronunciations of a word were written as separate words in the dictionary [12,13,14].

- Decoding Engine

The decoding engine searches for the best phoneme sequence given the feature and the model. For this HMM based system, the Viterbi decoding is used as the decoding engine.

Hidden Markov Model models spoken utterances as the outputs of finite state machines (FSMs). In this paper we experimented using left-to-right model of HMM which allows states to transit to themselves or to successive states but restricts transition to earlier states. Figure 1 represents left-to-right model of three emitting states of HMM

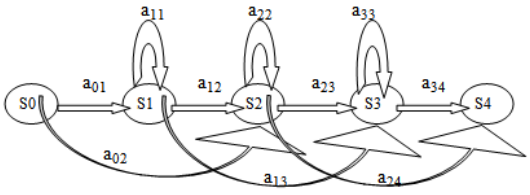

Figure 1. The topology of the HMM.

The most probable word Ŵ **is** given the observation sequence *O* that can be computed by taking the product of two probabilities for each word, and choosing the word for which this product is greatest. The components of the HMMs speech recognizer which compute those two terms are; the P(W), the prior probability computed by the language model and the P(O|W), the observation likelihood is computed by the acoustic model.

$$\hat{W} = \arg\max_{W \in L} \overbrace{P(O \mid W)}^{\text{likelihood}} \overbrace{P(W)}^{\text{prior}} \quad (1)$$

where Ŵ = the most probable word
O = observation sequence
W = word sequence
The architecture of Oriya isolated word recognition based on HMM is shown in the Figure 2

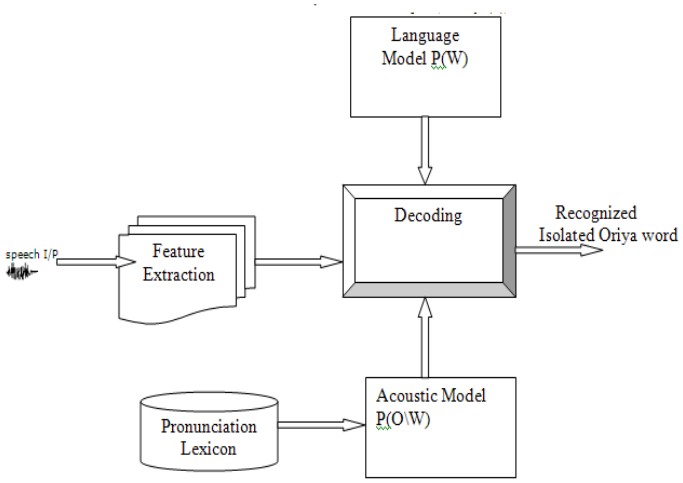

Figure 2. Architecture of Oriya isolated word recognition based on HMM

5. EXPERIMENTAL RESULTS

The performance of the Oriya isolated word recognition system was measured by computing the recognition accuracy at the word level. The decoding problem is solved using the Viterbi algorithm. The system was tested on 5 speakers on both seen and unseen data. Table 2 shows the performance of Oriya isolated speech recognition system. The first row refers to the case when the system is fed with the speech data with which the system was trained i.e. seen data. The second row indicates to the test data, i.e., the speech data that the system has not "seen" during training. The word accuracy is defined as

$$\text{word accuracy} = \frac{100 * \text{No. of correctly recognized words}}{\text{Total no. of words in the test suite}}$$

Table2. Performance of Oriya isolated speech recognition system for training and test data.

Types of data	***Word accuracy (%)***
Seen data	76.23
Unseen data	58.86

6. CONCLUSION AND DISCUSSION

This research paper tried to focus on problems faced by visual impaired students while answering in the examination and proposed a solution by implementing isolated Oriya recognition system based on left-to-right HMM which will be definitely helpful to provide fairness of assessment to the Oriya speaking visual impaired students. A baseline isolated Oriya speech recognition

was developed and the result is pretty encouraging. For actual deployment in appropriate task domain needs more performance accuracy rate by incorporating more training data as well detail modeling of speech signal. This work will be extended to the further research on connected word recognition as an extension of isolated speech recognition. The current system employs MFCC to represent speech frames. This feature vector does not take into account the systematic temporal variation of spectrum. The accuracy of the recognition system can be improved by using dynamic features such as delta-MFCC. Feature sets other than MFCCs can be tried such as Linea Predictive Coefficients (LPCs), Linear Predictive Cepstra (LPCepstra), log-scaled filterbank energies (FBANK) and PerceptualLinear-predictive coefficients (PLPs). The noise suppression techniques can be used to improve performance of speech recognition system as the training and test data contain noise which reduces the performance of the recognition system.

ACKNOWLEDGEMENTS
Financial support of DIT, MCIT Govt. of India for this work is highly acknowledged.

7. REFERENCES

[1] http://www.censusindia.net

[2] Samudravijaya K. ,Smitha Nair, Minette D'lima, "Recognition of spoken number", Proc.sixth Int. workshop on Recent Trends in speech, Music and Allied Signal Processing, New Delhi, 2001,pp. 1-5.

[3] Rabiner, L.R., Schafer, R.W, Digital Processing of Speech Signals, **P**earson education, 1st Edition,2004

[4] Quatieri T.F., Discrete-Time Speech Signal Processing, Pearson education ,Third Impression-2007.

[5] "CMU sphinx - open source speech recognition engines" htttp://www.speech.cs.cmu.edu/sphinx/.

[6] Lawrence Rabiner and Biing-Hwang Juang. Fundamentals of SpeechRecognition. Englewood Cliffs, NJ: Prentice Hall, pages 333-352 and 434-450,1993.

[7] Becchetti,C, Ricotti,L,P, Speech Recognition Theory and C++ Implementation, Jhon Wiley & Sons,2009.

[8] Samudravijaya, K., "Hindi Speech Recognition", J. Ac. Soc.India, vol.29,no.1,2001,pp.385-393

[9] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speechrecognition. pages 256–286. Proceedings of the IEEE, 1989.

[10] L. R. Bahl, S.V. De Gennaro, P.S.Gopalakrishnan, R.L. Mercer, "A fast approximate acoustic match for large vocabulary speech recognition", IEEE Transactions on Speech and Audio Processing, Vol. 1 Issue. 1, 59 –67, Jan.1993.

[11] Samudravijaya K and Maria Barot. A comparison of public domain software tools for speech recognition. pages 125–131. Workshop on Spoken Language Processing, 2003.

[12] Carnegie Mellon University. CMU pronuncing dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

[13] Wessel, F., Schluter, R. Macherey, K. and Ney, H., "Confidence Measures for Large Vocabulary Continuous Speech Recognition", IEEE Trans. on Speech and Audio Proc., 9(3):288–298, 2001

[14] Somnath Majumder and A.K.Dutta, ``Automatic recognition of isolated Bengali words'', Proc. Int. Workshop titled ``Speech technology for man-machine interaction'', Eds. PVS Rao and BB Kalia, Tata McGraw Hill, 1990, pp. 283-288.