International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Automatic Speech Recognition (ASR) System for Isolated Marathi Words: Using HTK

Sunil B. Patil, Nita V. Patil, Ajay S. Patil

Abstract*: The present manuscript focuses on building automatic speech recognition (ASR) system for Marathi language (M-ASR) using Hidden Markov Model Toolkit (HTK). The M-ASR system gives the detail about experimentation and implementation using the HTK Toolkit. In this work total 106 speaker independent Marathi isolated words were recognized. These unique Marathi words are used to train and evaluate M-ASR system. The speech corpus (database) is created by us using isolated Marathi words uttered with mixed gender people. The system uses Mel Frequency Cepstral Coefficient (MFCC) for the purpose of extracting features using Gaussian mixture model (GMM). Viterbi algorithm based on token passing is used for decoding to recognize unknown utterances. The proposed M-ASR system is speaker independent. The proposed system has reported 96.23% word level recognition accuracy.

Keywords* : Automatic Speech Recognition, Marathi, Hidden Markov Model, HTK, Isolated Words, Mel-Frequency Cepstral Coefficient.

I. INTRODUCTION

The current development in field of sciences and technology has resulted into immense need of communication between humans and machines [1]. Speech can be used as an effective media for communication between humans and computers. Most of the research in automatic speech recognition (ASR) has been carried out and reported for English and Western languages like French, German etc. [2]. ASR system has been developed for Indian languages such as Hindi, Assamese, Tamil, Telugu, and Bangali etc. Little work is reported for standard Marathi language. The Marathi is official and commonly spoken language of Maharashtra state in India. Multiple dialects of Marathi are spoken in Maharashtra, and also by millions of Marathi speakers who have migrated to other states in India and other parts of the world. It has become a necessity to construct ASR system for Marathi language.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
Sunil B. Patil, School of Computer Sciences(SOCS), Kavayitri Bahinabai Chaudhari North Maharashtra University(KBCNMU), Jalgaon, India Email: spatil512@gmail.com
Nita V. Patil, School of Computer Sciences(SOCS), Kavayitri Bahinabai Chaudhari North Maharashtra University(KBCNMU), Jalgaon, India Email: nitaapatil@gmail.com
Ajay S. Patil*, School of Computer Sciences(SOCS), Kavayitri Bahinabai Chaudhari North Maharashtra University(KBCNMU), Jalgaon, India Email: ajaypatil.nmu@gmail.com

In proposed research we have developed automatic speech recognition system for Marathi using HTK toolkit based on hidden markov model (HMM).Major tasks required to implement ASR system is speech corpus development, feature extraction, defining HMMs, training HMMs and speech recognition are shown in figure 1.

The paper is organized into six sections. The first section explores introduction and need of ASR for Marathi language. The work related to speech recognition in other Indian languages is described in second section.

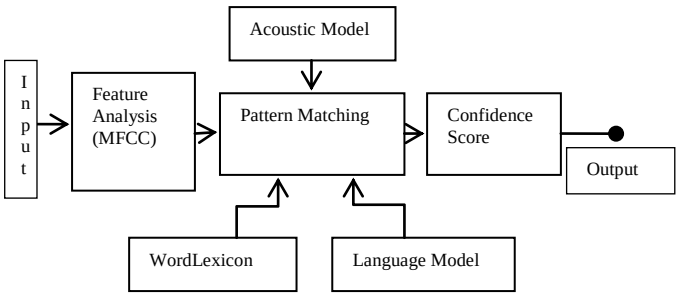

Fig. 1 . Block diagram of automatic speech recognition system.

Third section describes the process of data collection and corpus creation. Forth section describes building acoustic model of ASR for Marathi. Fifth section focuses on results and evaluation of the system followed by conclusion and future scope in sixth sections.

II. RELATED WORK

To design automatic speech recognition system many researchers have used HTK. Samudravijaya K. proposed recognition system for Hindi language uses context independent acoustic phonetic features and reported 93% recognition accuracy [4]. Kumar K. et al. have implemented ASR system for Hindi language using HTK with connected words, uttered by twelve different speakers. The system has reported 87.01% performance [2]. Dua M. et al. presented Punjabi automatic speech recognition system using HTK. The system is trained and tested using 115 distinct words collected from eight speakers. The system was evaluated for both trained and untrained environments with recognition performance 95.63% [3]. Akila A. et al. developed ASR system for Tamil language, with vocabulary size of 10 words. Dataset is gathered from two speakers. Spectral features have been extracted from each word. Word error rate reported by the system is 90% [5]. Sneha V. et al. implemented speech recognition system for Kannada languages with dataset 10 words collected from 14 speakers. The authors have developed speaker independent system using spectral and prosodic features with 90% recognition accuracy [6].

Mukundan S. et al. established ASR for Malayalam language using HTK toolkit. The database used contains 210 isolated spoken words. For training and testing purpose dataset was categorized into three sections. The system was tested on trained and untrained dataset environment. The recognition accuracy reported for this system in trained dataset mode varies from 84% to 88% and for untrained dataset mode varies from 94% to 100% [7]. Paulose S. et al. proposed ASR for Marathi using Kaladi toolkit. Three fold validation test conducted with deep neural network. Word error rate of the reported system is 21.2% [8]. Patil A. had reported the speaker independent ASR system using (HTK) for Ahirani language. Ahirani is one dialect of Marathi language. Dataset consisting of 20 isolated words has been used. Accuracy of recognition reported by the system is 94% [9]. Gawali B. et al. developed ASR for Marathi using 175 distinct sentences collected from 35 speakers both male and female. They have used spectral features as MFCC and dynamic time wrapping (DTW) for recognition. The systems have reported accuracy for MFCC 94.65% and accuracy for DTW is 73.25% [10]. Ghule K. et al. presented ASR for Marathi using neural network. Discrete wavelet transforms (DTW) is used as feature extraction purpose. Performance of the system is reported 60% [11].

III. DATA ACQUISITION

Dataset collection and preparation for Marathi ASR system is important task. Since this type of database is not available. We have built our own corpus which compromises total 80 sentences, from which we extracted 106 unique isolated words. The following section discusses the database preparation method.

A. Database Preparation

In the proposed ASR system 107 models are trained using unique isolated words, out of which one model is used for "sil" purpose. Recording is done with Audacity 2.3.0 [12] cross platform sound editor toolbox with the help Sennheiser-PC-350 built in micro headphone. The sampling frequency is set to 16000 kHz. Each word is recorded seven times from three speakers. Five samples are used to train the model, remaining two are used to test the model. Words are segmented manually for labeling purpose using Wavesurfer-1.8.5 [13]. Acoustical testing is carried out on developed database for recognition purpose. HCopy is used to extract MFCC features. HCopy tool is processed with MFCC_0_D_A (39-coefficient) configuration parameter and other important configuration parameters are shown in table1

Table-I: configuration parameters

Sr.No.	Parameters	value
1	Input file	.wav
2	Sample rate	16000 khz
3	Window size	250,000.0
4	Use Hamming	true
5	Filter bank used	26
6	Feature Parameter	MFCC_0_D_A
7	MFCC coefficient	12
8	Preemcoef	0.97
9	cepstral liftering used	22
10	Energy normalisation	True

IV. BUILDING OF THE MODEL WITH HMM

Acoustic Models are built in two ways, word model and phoneme model. In the proposed system word model described at url "http://www.labunix.uqam.ca/~boukadoum_m/DIC9315/Notes/Markov/HTK_basic_tutorial.pdf" is considered.

A. Model Initialization

The word models in proposed system are initialized using HMM Proto. The number of states and details of each state such as number, mean, variance etc. defined by HMM Proto is shown below,

```
~o <VecSize> 39 <MFCC_0_D_A>
~h "aahe"
<BeginHMM>
<NumStates>6
<State>2
<Mean>39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance>39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State>3
<Mean> 39
0.0 0 .0.(…) 0.0
<Variance> 39
1.0 1.0 (…) 1.0
<State>4
<Mean> 39
0.0 0 .0 (…) 0.0
<Variance> 39
1.0 1.0 (…) 1.0
<State>5
<Mean> 39
0.0 0.0.(…) 0.0
<Variance>39
1.0 1.0(…)1.0
<TransP>6
0.00.50.50.00.00.0
0.00.40.30.30.00.0
0.00.00.40.30.30.0
0.00.00.00.40.30.3
0.00.00.00.00.50.5
0.00.00.00.00.00.0
<EndHMM>
```

In proposed M-ASR system 107 HMM's Proto's are defined using above structure process for further processing. HInit tool is used to initialize each word model using HMM which uses modeled acoustical event (.mfcc) and speech label file (.lab). Topology used in this system consists of total six states [four states are active, two are non-emitting: first and last with no observation] with single Gaussian distribution with diagonal matrices. After initializing word model, it is then trained with HRest tool for estimating optimal values. This process has to be repeated until estimated optimal values do not converge. The process is known as re-estimating the model. Proposed ASR system uses the similar methodology described by Kuldeep Kumar et al. [2] in which word model is re-estimated three times for every word model.

B. Task Definition

Recognition performance of the system relies on the task grammar and task dictionary, both are written in text file format which is provided by HTK. The task grammar is executed with HParse tool to create word network model in standard lattice format (.SLF) for all defined HMMs. The dictionary (pronunciation model) consists of words along with their corresponding phonemes in a specified format Extended Backus-Naur form accepted by HTK. The task dictionary is compiled with HSgen tool to check the grammar for correctness [14]. The performance M-ASR (Marathi) depends on network (.slf) model.

V. RECOGNITION OF MARATHI ASR

Recognition performance of the M-ASR (Marathi) is processed with HVite tool [5]. HVite tool is used to recognize particular Marathi word from unknown utterance transcription.

A. System Testing

This is the final stage for the recognition of unknown utterances. Using HVite tool, M-ASR will identify a speech signal which is matches it against a word recognizer's of all HMMs and display recognition transcription in the form of rec.mlf format shown in figure 2. HVite tool is uses the viterbi algorithm which based on token. HVite tool receive input viz. HMM definition, marathi word dictionary, task grammar, hmm list, net.slf. In the proposed system, 107 Physical and Logical HMMs are observed with 112 lattice node and 220 arcs are observed for the recognition purpose. Figure 2 shows recognition accuracy of the isolated word model.

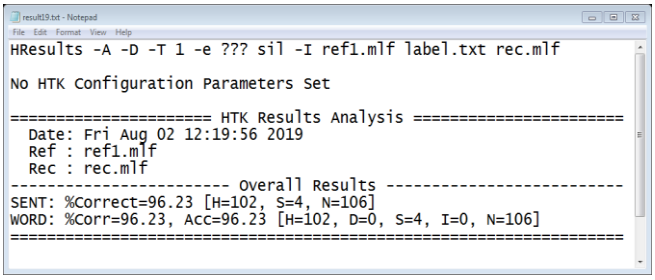

Fig. 2. Recognition accuracy of the isolated word model

B. Performance Evaluation

The performance of the proposed system is measured by computing recognition accuracy at word level. The system performance is analyzed using HResult tool. Table 2 shows the recognition accuracy of the proposed system. Here, Recognition accuracy is measured by using equation (1), Percentage Accuracy is measured by equation (2) and Word Error Rate is calculated using equation (3) Where N is the number of words given for testing, D is number of deletions S is number of substitutions and I is used for insertions.

$$\text{Recognition Accuracy}(\text{R.A.}) = \frac{[\text{N}-\text{D}-\text{S}]}{N} \times 100 \quad \ldots\ldots\ldots (1)$$

$$\text{Percentage Accuracy}(\text{P.A.}) = \frac{[\text{N}-\text{D}-\text{S}-\text{I}]}{\text{N}} \times 100 \ldots\ldots\ldots (2)$$

$$\text{Word Error Rate}(\text{W.E.R.}) = 100 - \text{P.A.} \quad \ldots\ldots\ldots\ldots (3)$$

Equation 1, 2 and 3 is used to estimate the sentences recognition accuracy, recognition performance of the word model and errors in word model of the proposed M-ASR system respectively.

Table- II: Recognition performance of Marathi ASR

Recognition Accuracy				
Number of spoken words for testing	*Number recognized spoken words*	*R.A.*	*P.A.*	*W.E.R.*
		Recognition accuracy	*Percentage accuracy*	*Word error rate*
106	102	96.23	96.23	3.77

VI. CONCLUSION AND FUTURE WORK

The present manuscript describes detailed study of M-ASR system for isolated word using HTK. The small scale standard database for (80 sentences & 106 words) unique isolated words for Marathi language is successfully developed. Data was segmented and labeled for training and testing purpose. The HTK tools HCopy, HInit, HCompv, HRest, Hparse, and HSgen were used for training and testing of the system. The system performance was analyzed using HResult. The proposed system has given recognition accuracy of 96.23% with very low word error rate 3.77%. The system is found to be efficient for Marathi language. The work implemented in this system is first step towards large scale M-ASR systems. In future, the work can be extended to connected words speech recognition system for large dataset.

REFERENCES

1. Nicolas Moreau, HTK(v3.1): Basic Tutorial downloaded: http://www.labunix.uqam.ca/~boukadoum_m/DIC9315/Notes/Markov/HTK_basic_tutorial.pdf
2. Kumar K., Aggarwal R., Jain A., "A Hindi speech recognition system for connected words using HTK." *Int. Journal of Computational Systems Engineering*1(1) 2012, pp. 25-32.
3. Dua M., Aggarwal R., Kadyan V., Dua S., "Punjabi automatic speech recognition using HTK.", *Int. Journal of Computer Science Issues* 9(4) 2012, pp. 359-364.
4. Samudravijaya K., "Computer recognition of spoken Hindi." *In Proc. of Int. Conference of Speech, Music and Allied Signal Processing, Triruvananthapuram*, 2000, pp. 8-13.
5. Akila A., E. Chandra., "Isolated Tamil word speech recognition system using HTK.", *Int. Journal of Computer Science Research and Application* 3(2) 2013, pp. 30-38.
6. Sneha V., Hardhika G., Jeeva K., Gupta D., "Isolated Kannada Speech Recognition Using HTK- A Detailed Approach.", *In Progress in Advanced Computing and Intelligent Engineering, Springer, Singapore* 2018, pp. 185-194.
7. Mukundan S., Bhasha S., "Malayalam Speech Recognition using HTK.", *Int. Journal of Advanced Computing and Communication Systems*1(1) 2014, pp. 1-5
8. Paulose S., Nath S.,Samudravijaya K. , "Marathi Speech Recognition", *Proc. of The 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages*, pp. 230-233, 29-31 August 2018, Gurugram, India
9. Patil A., "Automatic Speech Recognition for Ahirani Language Using Hidden Markov Model Toolkit (HTK)." *Int. Journal of Computer Science Trends and Technology* 2(3) 2014, pp.140-144.
10. Gawali B., Gaikwad S., Yannawar P., Mehrotra S., "Marathi isolated word recognition system using MFCC and DTW features." *ACEEE Int. Journal on Information Technology* 1(1) 2011, pp. 21-24.
11. Ghule K., Deshmukh R., "Automatic Speech Recognition of Marathi isolated words using Neural Network.," *Int. Journal of Computer Science and Information Technologies* 6(5) 2015, pp. 4296-4298
12. https://www.audacityteam.org/.

13. Speech, Music and Hearing part of School of Computer Science and Communication: https://www.speech.kth.se/wavesurfer/
14. Choudhary A., Chauhan R., Gupta G., "Automatic Speech Recognition System for Isolated and Connected Words of Hindi Language By Using Hidden Markov Model Toolkit (HTK)."In Proceedings of Int. Conference on Emerging Trends in Engineering and Technology, 2013, pp. 847-853.

AUTHORS PROFILE

Sunil B. Patil is M.Sc. in Computer Science and pursuing Ph.D. in Computer Science from School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon (M.S.) India. He is currently working on speech recognition. He is actively participating in research competitions, conferences, seminars and workshops etc. He has 8 years of teaching and more than 5 years of research experience. His research areas include - Machine Learning, Neural Network and Speech Recognition. He has Programming experience in Java, Tomcat Apache, PHP and Assembly Language.

Nita V. Patil is Doctorate in Information Technology from Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, Maharashtra. She has completed her Master's Degree in Information Technology from the same University. She is working as an Assistant Professor in School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, Maharashtra. She has total teaching experience of 19 years with more than 20 publications in reputed, peer reviewed National and International Journals, Books & Conferences. Her research area includes- Information Extraction, Natural Language Processing and Machine Learning. She has successfully carried out two research projects funded by funding agencies like Rajiv Gandhi Science and Technology Commission (RGS&TC), Govt. of Maharashtra and KBCNMU, Jalgaon etc. She is also a member of International Association of Engineers.

Ajay S. Patil is Doctorate in Computer Science from North Maharashtra University, Jalgaon, Maharashtra. He has completed his Master's Degree in Computer Science from the same University. He is working as a Professor and Head, Department of Computer Applications, School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, Maharashtra. He has total teaching experience of 21 years with more than 55 publications in reputed, peer reviewed National and International Journals, Books & Conferences. He has successfully guided 3 M.Phil. and 4 Ph.D. students. His research area includes- Information Retrieval, Natural Language Processing and Machine Learning. He has successfully carried out four research projects funded by funding agencies like UGC, New Delhi, Rajiv Gandhi Science and Technology Commission (RGS&TC), Govt. of Maharashtra etc. He has also worked as a session chair and resource person for workshops, conferences, seminars at national and international level and delivered expert lectures at refresher and orientations programs. He is also a member of various National and International professional societies in the field of engineering & research like Member of IAENG (International Association of Engineers)