International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-3, February 2020

Speakar-Independent Isolated Word Recognition using HTK for Varhadi – a Dialect of Marathi

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

Abstract*: Speech recognition is widely used in the computer science to make well-organized communication between humans and computers. This paper addresses the problem of speech recognition for Varhadi, the regional language of the state of Maharashtra in India. Varhadi is widely spoken in Maharashtra state especially in Vidharbh region. Viterbi algorithm is used to recognize unknown words using Hidden Markov Model (HMM). The dataset is developed to train the system consists of 83 isolated Varhadi words. A Mel frequency cepstral coefficient (MFCCs) is used as feature extraction to perform the acoustical analysis of speech signal. Word model is implemented in speaker independent mode for the proposed varhadi automatic speech recognition system (V-ASR). The training and test dataset consist of isolated words uttered by 8 native speakers of Varhadi language. The V-ASR system has recognized the Varhadi words satisfactorily with 92.77%. recognition performance.

Keywords: Speech Recognition (SP), Varhadi, HMM, HTK, Isolated Words, Mel Frequency Cepstral Coefficient (MFCCs), PLP, Speaker Independent, Interactive Voice Response (IVR).

I. INTRODUCTION

Speech signal is the most important way to make efficient communication between humans and computers. It is very easy to make communication via speech rather than keyboard and mouse. It is very beneficial to those who are semi-literate. The motivation behind this work is to reduce the gap between humans and computers. Very few researches is reported related to Varhadi dialect of the Marathi language [1]. The field of speech recognition has achieved tremendous progress for European languages as compared to Indian regional languages. In India several researchers have worked together for development of ASR system such Assamese, Bengali, Kannada, Hindi, Manipuri, Oriya, Marathi, Tamil, Punjabi, and Telugu [2]. Varhadi is a spoken language of Vidharbha region. Vidharbha is the eastern region of the Indian state of Maharashtra; it constitutes Nagpur, Amravati and Akola division. A majority of Vidharbhians speak Varhadi, Dangi and Zadi dialects of Marathi [3]. This work is an effort for every semi-literate person in Maharashtra which will help them to communicate with computer in their regional language. For Indian regional spoken languages work is not yet reached to apprehensive level [4]. So there is wider scope to develop speech recognition system for Varhadi dialect of Marathi languages. The paper is organized into seven sections. Sections I illustrates the importance of Varhadi language ASR, Section II gives the literature survey related to Indian regional languages, Section III focuses on feature extraction using HMM, Section IV is related to the data collection, detailed study of the model generation is corporate in Section V. section VI is related to the results analysis and Section VII concludes the paper with future scope of work.

Revised Manuscript Received on February 05, 2020.

* Correspondence Author

Sunil B. Patil*, Research scholar, Department of Computer Science, KBCNMU, Jalgaon, India. E-mail: spatil512@gmail.com

Nita V. Patil, Assistant Professor, Department of Computer Science, KBCNMU, Jalgaon, India. E-mail: nitaapatil@gmail.com

Ajay S. Patil, Professor, Department of Computer Science, KBCNMU, Jalgaon, India. E-mail: ajaypatil.nmu@gmail.com

II. RELATED WORK

This section presents literature review of some research related to ASR system for Indian languages. According to the technical report presented by Kamper H. et al. a large amount work has been reported in the field of speech technology for English and European languages [5]. The researchers in this field have mainly focused on isolated spoken words, digit and alphabets [6]. Bharali S. et al. [7] have developed a digit recognizer for Assamese using fifteen speakers including male and female. Ten words from each speaker are recorded ten times. Total database consists of 1500 words/samples. MFCC parameter on clean database is used in the experiment and 80% recognition accuracy is reported. Shahanwazuddin S. et al. [8] have proposed a system that accesses the price of agriculture commodities using interactive voice response (IVR) and ASR modules. MFCC parameter on the database consisting of 138 isolated words used for features extraction. 8% performance improvement in baseline is reported. Banerjee P. et al. [9] presented a system for Bengali language using triphone clustering in acoustic modeling. 4000 utterances from 22 speakers are recorded (18 males and 4 females respectively). 76.33%. average recognition accuracy using MFCC parameter is reported. The ASR system using HTK presented by Das B. et al. [10] consist of 19640 unique words recorded from 70 male and 40 female speakers in Bengali. Authors have used MFCC and LPCC feature extraction parameters. Recognition accuracy reported is 85.3% (using MFCC) and 79.6% (using LPCC). Authors have proved that recognition rate is more using MFCC features than LPCC. Choudhary et al. [4] have presented an approach that uses statistical algorithm to develop the ASR system for Hindi language. 100 distinct isolated words are considered in the experiment. 95% overall accuracy is reported for 100 distinct isolated words. Kumar A. et al. [11] have design speaker dependent Hindi speech recognition system for 10 unique isolated utterances. MFCC and Perceptual Linear Predictive (PLP) has been used for features extraction.

The speech recognition system has reported 95.40% accuracy. Bhardwaj I. et al. [12] have presented the speaker dependent and independent systems using K-Means algorithm using 10 Hindi words. 99% recognition rate of speaker dependent system and 98% recognition rate of speaker independent system that have used multiple speakers is reported. Aggarwal R. et al. [13] have proposed statistical pattern classifier model for Hindi languages. The dataset is divided into four parts consisting of 100, 150, 200 and 250 isolated words respectively. Recognition accuracy reported for each dataset is 78%, 72%, 65% and 56% respectively. Kumar A. et al. [14] have presented continuous Hindi speech recognition system using GMM and achieved recognition accuracy 97.04%. Patil P. et al. [15] have built the speaker dependent system using Bi-gram language model. Dataset used in the system includes 10 Marathi sentences. Authors have reported 85.65% overall recognition accuracy. Kamble V. et al. [16] have presented emotion recognition system for Marathi Language. The system recognizes emotions such as happy, sad, angry etc. Seven isolated words are included in the dataset developed for the emotion recognition system. The system has reported 83.33% emotion recognition accuracy. Patil A. [17]., have built Ahirani ASR system using HMMs for Marathi dialect of the Ahirani language. speech Database is consist of 20 isolated words. The recognition rate of the Ahirani ASR system is 94%. Bansod N. et al. [1] have proposed the system for Varhadi language which is also another dialect of Marathi language. MFCC and LPC features extraction techniques have been used for speaker recognition. Recognition accuracy reported using MFCC is 60.33% and using LPC is 85%. Mehta L et al. [18] have presented ASR system for Marathi language using MFCC and LPCC features extraction techniques. 48 isolated words are used to train the system. The system which has extracted features using MFCC has reported 99% accuracy and the use LPCC to extract features has reported 77% accuracy. Gandhe et al. [19] have reported isolated word recognition system for Marathi language that combines features extracted by the MFCC and DTW techniques for online and offline speakers. The system has reported maximum 100% accuracy for offline and 72.22% for online dependent speakers. Dua et al. [20] have built a Punjabi isolated speech recognition system that has used HTK. The system has used MFCC technique to train the model and reported 95.63% recognition accuracy.

III. ISOLATED VARHADI SPOKEN WORD RECOGNITION USING MFCC AND HTK

The literatures in the field of speech recognition have reported the use of feature extraction techniques such as LPC, PLP, LPCC and MFCC etc. for speech recognition. MFCC has proven to be more powerful feature extraction technique. Therefore, the ASR system presented in this paper also uses MFCC as a feature extraction technique. Mel-frequency cepstral coefficients (MFCC) are vectors of acoustical coefficients.

A. Features Extraction

The Mel-Frequency Cepstral Coefficient is power spectrum of voice, which is depends on linear cosine transform of log power spectrum on a nonlinear Mel Scale Frequency [21]. The optimized parameters for proposed system include overlapping frames of 25 millisecond duration. The shift between successive frames being 10 frames and multiplied by a hamming window. 12 cepstral coefficients were derived from the output of 26 Mel filters. The cepstral coefficients were liftered using a raised sine window length 24. The 12 liftered mel-frequency cepstral coefficients (MFCC) from the feature vector representing a frame of speech [24]. The process of features extraction is shown in Fig.1.

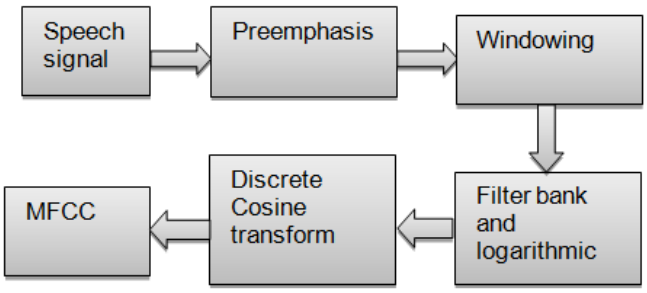

Fig. 1. Feature extraction process of MFCC

B. Hidden Markov Model Toolkit (HTK)

The Hidden Markov Model (HTK) is developed by Cambridge University Engineering Department CUED). This software is freely available on registration [22]. Now a days HTK is widely used in the field of mobile application, IVR application, and pattern recognition and especially in the field of speech recognition [23].

IV. DATA COLLECTION FOR VARHADI LANGUAGE

The prime interest in the present work is to develop standard database for Varhadi language. The speaker independent named as V-ASR system is trained with words uttered by 8 native speakers of different age groups. This system includes 83 isolated varhadi words with 1170 speech file. Each word is recorded in Audacity 1.3 Beta Unicode sound editor toolbox and further manually labeled in wavesurfer-1.8.8p5 toolbox. The Sennheiser PC-350 with built in microphone is used for recording purpose at a sampling rate of 16MHz. The details regarding Varhadi unique isolated word database are shown in table I.

Table I: Varhadi Isolated Words

अंन	शब्द	अंन	शब्द
१	आलता	२	आमच्याईकडे
३	आमी	४	अनं
५	आन्ली	६	अय
७	अलप	८	आताच
९	आठठाईस	१०	चलतं
११	चालंला	१२	चाल्ले
१३	चंनचंन	१४	दाखोलं
१५	दवाले	१६	देल्ला
१७	देल्ले	१८	देसिन
१९	देतं	२०	ढोर
२१	दिलं	२२	गावले

२३	गेलतो	२४	गेल्ते
२५	हल्ली	२६	हाव
२७	हाय	२८	हाय्यले
२९	हिच्च्या	३०	इकत
३१	इकायला	३२	इलें
३३	इतकुशच	३४	जावून
३५	जेच्याच्यान	३६	जेवला
३७	कावून	३८	कायले
३९	कायचे	४०	खाल्ला
४१	खायले	४२	कोटा
४३	लागुन	४४	लई
४५	लेका	४६	मले
४७	म्हासाड	४८	माया
४९	म्या	५०	म्यान
५१	नेलते	५२	पाजलं
५३	पतलं	५४	प्याला
५५	पेले	५६	पेली
५७	राबतो	५८	रायलां
५९	रायली	६०	रोजनं
६१	सकाई	६२	सोकुन
६३	तैईचा	६४	टाकलं
६५	टावेल	६६	तेनं
६७	तिनं	६८	तुले
६९	तुमाले	७०	तुमच्याइकडे
७१	तुया	७२	तुयाकडे
७३	त्यो	७४	त्यानं
७५	त्याहिले	७६	त्याईनं
७७	व्हतं	७८	व्हती
७९	वावरत	८०	यायले
८१	येचलां	८२	येवून
८३	झंझट		

V. ACOUSTICAL MODEL AND TASK GRAMMAR

An acoustical word model is used as a reference model to compare unknown utterances using Viterbi algorithm. Two types of acoustical models are defined word model and phoneme model. Word model is initialized using HMM and is implemented by defining HMM Proto [25].

A. Acoustical Model Generation

Define 84 HMM Protos are used in the present system. The feature vector is based on 13 mel-cepstral coefficients (MFCC) and their time derivative acoustic phones were modeled by HMM with six emitting states. The distribution of features for each state was modeled by Gaussian Mixture Model (GMM). The states used in HMM Proto are given in Fig. 2

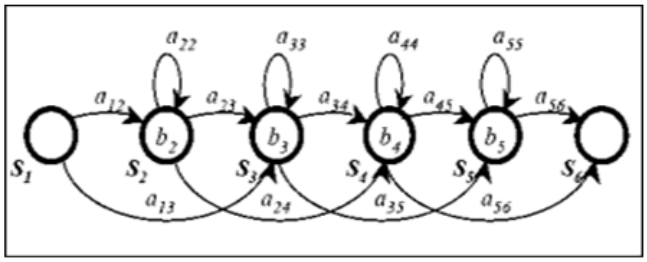

Fig.2 HMM topology using GMM for each word

In the training phase HTK tool is used to estimate optimal values from HMM proto (transition probability, mean, and variance vectors for each observation function. This process has to repeat as per the need. In present system 3 iteration are considered, hence three HMM per word in dataset needs to be generated.

B. Task Grammar and Dictionary of V-ASR

It is necessary to define task grammar and word dictionary (pronunciation dictionary). In V-ASR (Varhadi), task grammar and Varhadi word dictionary are defined as per Extended Back-Naur Form (EBNF) which is written in text file. HParse tool is used to generate network model (.slf). The names of the labels are also added to the correspondence to represent the symbols that will be output of by the recognizer.

C. Testing of V-ASR

The final stage in the system is generating transcriptions of unknown utterance is testing [11]. In this stage testing signal are converted into series of acoustical vectors (.mfcc) using HTK tool HCopy. Input observations along with HMM definitions, Varhadi word dictionary, task network(.slf), and names of HMMs defined in HMMs list is taken by HTK tool HVite to generate output transcription(.mlf) file. The testing is done using HVite tool which process the speech signal with Viterbi algorithm to compare the test utterances with reference transcriptions defined into word dictionary.

VI. RESULT ANALYSIS

The performance of the V-ASR system is measured at the word level [24]. The system performance is analyzed using HTK tool HResult. Figure 3 and table 2 shows the word correction rate of the V-ASR system. The word correction rate (WCR), word accuracy rate (WAR) and Word Error Rate (WER) is calculated using equation 1, 2 and 3 respectively [26].

$$Word\ Correction\ Rate = \frac{N - D - S}{N} * 100 \ldots eq.1$$

$$Word\ Accuracy\ Rate = \frac{N - D - S - I}{N} * 100 \ldots eq.2.$$

Where N is the number of words given for testing, S is number of substitutions, D is number of deletions and I is used for insertions. Equation (3) shows the calculation word error rate.

$$Word\ Error\ Rate(WER) = 100 - WAR \ldots . eq.3.$$

Equation (1) is used to calculate the word correction rate. Equation (2) is used to calculate word accuracy. Finally equation (3) is used to calculate word error rate

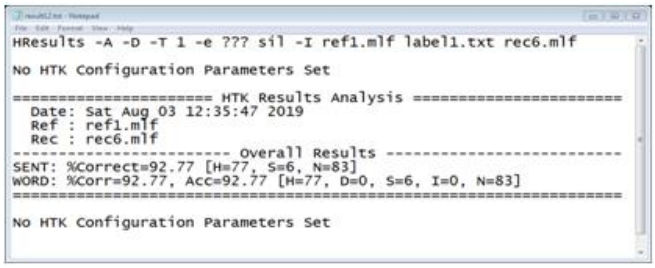

Fig. 3. Recognition accuracy of 83 Varhadi Isolated Words

Table II: Recognition Performance of V-ASR

Recognition Accuracy				
# spoken words for testing	*# recognized spoken words*	*W.C.R*	*W.A.R*	*W.E.R.*
		Recognition accuracy	*Percentage accuracy*	*Word error rate*
83	77	92.77	92.77	7.23

VII. CONCLUSION AND FUTURE WORK ANALYSIS

This paper has presented an automatic speech recognition system for Varhadi language. The proposed V-ASR system has successfully developed the standard data base particularly for Varhadi dialect of Marathi language. The system has achieved the 92.77%, recognition accuracy which satisfactory. The proposed system will be extended to large vocabulary in IVRS system for agriculture purpose.

REFERENCES

1. Bansod N., Dadhade S., Kawathekar S. Kale K., "Speaker Recognition using Marathi (Varhadi) Language", *Proceeding ICICA* 2014, pp 421-425
2. P.P. shrishrimal., R.R. Deshmukh., V.B. waghmare., "Indian language speech database-A review",. *Int. jour. of comp. application*. Vol 47, issue 5, pp-17-21. (2012)
3. Wikipedia contributors. Vidarbha [Internet]. Wikipedia, The Free Encyclopedia; 2019 Sep 6, 07:31 UTC [cited 2019 Sep 8]. Available from:https://en.wikipedia.org/w/index.php?title=Vidarbha&oldid=914266922
4. Annu A., M.R.Choudhary., M. G. Gupta,, "Automatic Speech Recognition System for Isolated and Connected Words of Hindi Language Using Hidden Markov Model Toolkit", *In Proceedings of Int. Conf. on Emerging Trends in Engineering and Technology*.2013, pp.- 847-853.
5. Kamper H, Niesler T., "A literature review of language, dialect and accent identification. Technical Report-1201", *Digital Signal Processing Lab.. Dept. of Electrical and Electronic Engg*, S.A. Stellenbosch University (2012)..
6. R. Djemili, M. Bedda, H. Bourouba, "Recognition of spoken arabic digits using neural predictive hidden markov models". *Int. Arab J. Inf. Technol.*, vol. 1, issue 2,(2004).pp. 226-233.
7. S. S. Bharali, S. K. Kalita, "A comparative study of different features for isolated spoken word recognition using HMM with reference to Assamese language", *International Journal of Speech Technology*, vol.18, issue4, (2015). pp. 673-684.
8. S. Shahnawazuddin, D.Thotappa, B. D. Sarma, A. Deka, S. R. Prasanna, R. Sinha,, "Assamese spoken query system to access the price of agricultural commodities", *IEEE, National Conference on Communications*, 2013, pp. 1-5.
9. P.Banerjee, G. Garg, P. Mitra, A. Basu,, "Application of triphone clustering in acoustic modeling for continuous speech recognition in Bengali", 19th *IEEE Int. Conference on Pattern Recognition*, 2008, pp. 1-4.
10. B. Das, S. Mandal, P.Mitra,, "Bengali speech corpus for continuous automatic speech recognition system", *International. Con. on speech database and assessments, IEEE* (2011), pp. 51-55.
11. A. Kuamr, M. Dua, A. Choudhary, "Implementation and performance evaluation of continuous Hindi speech recognition", *International. Con. on Electronics and Communication Systems, IEEE* 2014 pp. 1-5.
12. I. Bhardwaj, N. D. Londhe, "Hidden Markov Model based isolated Hindi word recognition", *2nd International. Conference on Power, Control and Embedded Systems, IEEE*, December 2012, pp.1-6.
13. R. K. Aggarwal, M. Dave, "Fitness Evaluation of Gaussian Mixtures in Hindi Speech Recognition System", *International. Conference on Integrated Intelligent Computing, IEEE August* 2010 pp.177-183.
14. A. Kuamr, M. Dua, T. Choudhary, "Continuous Hindi speech recognition using Gaussian mixture HMM", *Conference on Electrical, Electronics and Computer Science, IEEE*, March 2014, pp.1-5.
15. P. P.Patil, S. A. Pardeshi, "Marathi connected word speech recognition system, *International Conference on Networks & Soft Computing*, IEEE August 2014, pp.314-318.
16. V. V. Kamble, B. P. Gaikwad, D. M. Rana, "Spontaneous emotion recognition for Marathi Spoken Words", *International. Conference on Communications and Signal Processing*, April 2014 pp.1984-1990.
17. A.S. Patil, Automatic Speech Recognition for Ahirani Language Using Hidden Markov Model Toolkit (HTK). *International Journal of Computer Science Trends and Technology*, 2014, vol.2 issue 3, pp.-140-144.
18. L. R. Mehta, S. P. Mahajan, A. S. Dabhade, "Comparative study of MFCC and LPC for Marathi isolated word recognition system", *International. Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering.* vol. 2, issue 6, 2013, pp.2133-2139
19. M. B. Shinde, D. S. Gandhe, "Speech processing for isolated Marathi word recognition using MFCC and DTW features", *International. Jour. of Innov. in Engg. and Tech.*, vol.3, issue 1, 2013, pp.-109-114
20. M. Dua, R. K. Aggarwal, V. Kadyan, S. Dua,, "Punjabi automatic speech recognition using HTK", *International Jour. of Computer Science* vol. 9 issue 4 , 2012, pp.- 359.
21. S. B. Harisha, S. Amarappa, D. S. Sathyanarayana, "Automatic Speech Recognition-A Literature Survey on Indian languages and Ground Work for Isolated Kannada Digit Recognition using MFCC and ANN " *International. Journal of Electronics and Computer Science Engineering* vol.4, issue 1 , 2015, pp.91-105
22. Young, S., Gunnar, E., Gales M., Thomas H., Dan K., Xunying L., Gareth M., et al. The HTK book." Cambridge university engineering department 2000, 3: 175.
23. Huang, X., Li D., "An Overview of Modern Speech Recognition", 2010, pp.-339-366.
24. K. Samudravijaya, "Computer recognition of spoken Hindi", In Proceeding of International Conference of Speech Music and Allied Signal Processing Triruvananthapuram. 2000, pp. 8-13.
25. Nicolas Moreau, HTK(v3.1): Basic Tutorial downloaded:http://www.labunix.uqam.ca/~boukadoum_m/DIC9315/Notes/Markov/HTK_basic_tutorial.pdf.
26. B. A. Al-Qatab, R. N. Ainon, "Arabic speech recognition using hidden Markov model toolkit (HTK)", In Information Technology Vol.2, June 2010, pp. 557-562.

AUTHORS PROFILE

Sunil B. Patil is M.Sc. in Comp. Sci. and pursuing Ph.D. in Computer Science from KBC North Maharashtra University, Jalgaon (M.S.) India

Dr. Nita V. Patil is Ph.D. in Information Technology from KBC North Maharashtra University, Jalgaon (MS). She is working as an Asst. Professor in School of Computer Sciences, KBCNMU Jalgaon. She has total teaching experience of nineteen years with more than twenty publications in peer reviewed International and national Journals.

Dr. Ajay S. Patil is PhD in Computer Science from KBC North Maharashtra University, Jalgaon, (MS). He is Professor and Head, (Computer Applications), School of Computer Sciences, KBCNMU, Jalgaon, Maharashtra. His total teaching experience is twenty one years and has nearly sixty publications in peer reviewed national and international journals. He has successfully guided three M.Phil. & four Ph.D. students in the subject of Computer Science.