International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 07 | July-2016 www.irjet.net p-ISSN: 2395-0072

HTK Based Speech Recognition Systems for Indian Regional languages: A Review

Devyani S. Kulkarni[1], Ratnadeep R. Deshmukh[2], Pukhraj P. Shrishrimal[3], Swapnil D. Waghmare[4]

[1]*M.Tech Student, Dept. of Computer Science and IT, Dr. B.A.M.U, Aurangabad, [MS], India.*
[2]*Professor & Head , Dept. of Computer Science and IT, Dr. B.A.M.U, Aurangabad, [MS], India.*
[3, 4] *Research Student , Dept. of Computer Science and IT, Dr. B.A.M.U, Aurangabad, [MS], India.*

Abstract - *The Speech is most essential & primary mode of Communication among all human being. Human beings have long been motivated to create computer system that can understand and talk like humans. Speech Recognition is one of the important research areas. In regard of this the review of existing work on speech recognition is useful for carrying out further research. The aim of this paper is to give a brief overview of the Automatic Speech Recognition Systems which are built using HTK toolkit. This paper provides a literature survey of such systems which are built for recognizing Indian regional languages. Thirty papers are reviewed in this work from point of view of their Language, Year, Type of utterance, No. of Speakers, Utterances of each word, Recording environment, No. of words / sentences, Feature extraction technique used, Word accuracy, No. of states in HMM, Word error rate.*

Key Words: Automatic Speech Recognition (ASR), HTK Toolkit, Hidden Markov Model (HMM).

1. INTRODUCTION

The Speech is most essential & primary mode of Communication among all human beings [1]. Human beings have long been motivated to create computer system that can understand and talk like humans. In this direction, researchers are trying to develop systems which can analyse, classify and recognize the speech signals [2]. There are various spoken languages throughout the world. It is natural for people to expect speech interfaces with computer because Communication among the human being is dominated by spoken languages [3]. Also Human computer interaction (HCI) is crucial activity [4]. Interaction with machines through speech technology is easier instead of using other input devices like pointers and keyboard etc. The process of automatically recognition of speech by machines is known as Automatic Speech recognition. With the help of speech recognition technique it is possible to have an interaction between human beings and machine [5].ASR makes it possible for machines to convert a speech signals to the text format which is equivalent to the information conveyed by the spoken words with the help of identification and understanding process. ASR technology provides a natural pathway for communication between human being and machines also we can say this technology is the nothing but a key of technology of Man- Machine interface [6, 7].

The Automatic Speech Recognition (ASR) field of research is near about 60 years old. Many researchers have tried for advanced research and have successfully developed the particular systems. The first research work was carried out in the early era of 1950's at Bell Labs. And the research has been carried out up to what advanced technology we are using today.

The main objective of ASR systems is automatically obtaining the output string from input speech signals. This process can be illustrated in fig below [8].

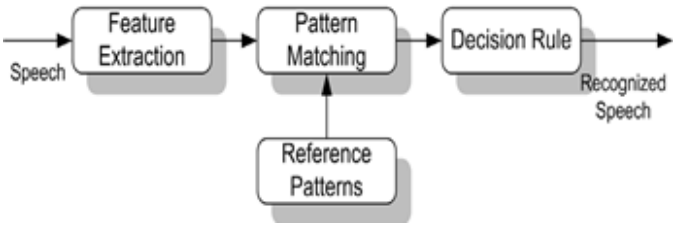

Fig. 1. Concept of ASR system [9]

This paper presents the research work in the development of Automatic speech recognition system using HTK toolkit for Indian languages.

The paper is organized as follows the section 2 explains what is meant by Hidden Markov Model and HTK Toolkit, section 3 describes Review of Speech Recognition Works done Using HTK Toolkit followed by table of comparison of different languages and conclusion.

2. HIDDEN MARKOV MODEL AND HTK TOOLKIT

When in system the future state of the system is depend on the present state of the system then this property of the system is called as Markov property. And the system which possesses such property is known as Markov process. When the Markov processes with hidden states are converted into statistical Markov model it is known as Hidden Markov Model [10].

Hidden Markov Model (HMM) is a doubly stochastic process with one that is not directly observable. This hidden stochastic process can be observed only through another set of stochastic processes that can produce the observation sequence. HMMs are the so far most widely used acoustic models. The reason is just it provides better performance

than other methods. HMMs are widely used for both training and recognition of speech system [11].

HMM are statistical frameworks, based on the Markov chain with unknown parameters. Hidden Markov Model is a system which consists of nodes representing hidden states. The nodes are interconnected by links which describes the conditional transition probabilities between the States. Each hidden state has an associated set of probabilities of emitting particular visible states [12].

The Hidden Markov Model Toolkit (HTK) is a portable toolkit for building and manipulating hidden Markov models. HTK is basically used for speech recognition research. Along with it has been used for numerous other applications including research into speech synthesis, character recognition and DNA sequencing. It is used worldwide. It contains liberty modules tools which are written in C language. HTK mainly run on UNIX platform but can also run on other platforms also. The tools provide sophisticated facilities for speech analysis, HMM training, testing and results analysis. The software supports HMMs using both continuous density mixture Gaussians and discrete distributions and can be used to build complex HMM systems. The HTK release contains extensive documentation and examples [13].

3. REVIEW OF SPEECH RECOGNITION WORKS USING HTK TOOLKIT

In this section we are going to present the review of speech recognition systems which were developed using HTK Toolkit for different languages.

3.1 Sanskrit Language

Jitendra Singh Pokhariya et al. (2014) [14] in their paper presented a work done for building a speech recognition system for Sanskrit language. The HTK toolkit was used for development of the particular system. The system was trained for recognizing 50 Sanskrit words. The database used was developed by taking speech samples from 10 speakers. The system showed the overall accuracy with 5 states of HMM topology as 95.2% and for 10 states 97.2%.

3.2 Ahirani Language

Patil A. S. (2014) [15] in his paper presented the implementation of HMM based speaker independent isolated word speech recognition system for Ahirani language. He have used HTK toolkit for developing the particular system which was trained with 20 Ahirani words. The database contains data recorded from 10 speakers. The system was tested by using data collected from another 10 speakers. The recording of database was done in room environment. The system has given 94 % accuracy.

3.3 Hindi Language

Shweta Tripathy et al (2013) in their paper have presented their work about developing the speech recognition system for Hindi language. For feature extraction various techniques like MFCC (Mel Frequency Cepstral Coefficient), LPC (Linear Predictive Coding) were used and HMM (Hidden Markov Model) was used as the classifier. For implementing this system authors have used HTK Toolkit. For sound recording audacity and Cygwin for executing the HTK commands in Linux type environment in windows platform. After testing the system in both speaker independent and speaker dependent environment performance result and comparison graph shows that the developed system gives good performance with MFCC as compared to LPC.

Annu Choudhary et al (2013) [16] have discussed the work of implementing the Speech Recognition system (ASR) for isolated words and connected words of Hindi language and working of HTK i.e. Hidden Markov Model Tool which is based on Hidden Markov Model (HMM) and also a statistical approach. Authors have used it for developing the proposed system. The system was trained for 100 distinct Hindi words initially. After testing the recognition results the system had shown the overall accuracy 95% and 90% for isolated words and connected words respectively.

Ankit kumar et al (2014) [17] in their paper they have compared the performance of continuous Hindi speech recognition system with different vocabulary sizes and feature extraction techniques. For feature extraction both Mel Frequency Cepstral Coefficient (MFCC) and Perceptual Linear Prediction (PLP) both are used. Hidden Markov Model (HMM) is used at back-end of an ASR system for monophone based acoustic modelling. System was implemented using HTK 3.4.1 toolkit. The 70 different Hindi words were used to train the system. At the end result shows that the system gave 95.08% accuracy for MFCC as a feature extraction technique.

Gaurav et al. (2012) [18] in their paper presented a work about application which is voice interface specific in local language for the benefit of technology to rural India. The goal they have set for this work was to make a continuous speech recognition system in Hindi for teaching Geometry in Primary schools. They have used the Mel Frequency Cepstral Coefficients for feature extraction and Hidden Markov Modeling for modeling the acoustic features. Hidden Markov Modeling Tool Kit –3.4 was used both for feature extraction and model generation. The Julius recognizer which is language independent was used for decoding. A speaker independent system is implemented and results are presented.

Preeti saini et al (2013) [19] have built a speech recognition system for Hindi language isolated words using Hidden Markov Model (HMMs). For recognizing isolated words they have used acoustic word model. For training the system 113 Hindi words were used. Data required for training had been collected from nine speakers. The result had shown that the accuracy of the particular system with 10 states in HMM was 95.49%.

Babita Saxena et al (2015) [20] in their paper presented a baseline digits speech recognizer for Hindi language. For the database they had collected the speech samples in different environment, so the recordings contain various noises like vehicle horns, door opening etc. After that

all these audio recorded data taken from 8 speakers was used to train the acoustic model. The vocabulary size of the recognizer was 10 words. Authors had used HTK toolkit for building acoustic model and evaluating the recognition rate of the recognizer. The efficiency of the recognizer developed on recorded data, is shown at the end of the paper and possible directions for future research work are suggested.

Kuldeep Kumar et al (2011) [21] have developed a speech recognition system for Hindi language. They have used Hidden Markov Model Toolkit (HTK) for developing the system. The system recognizes the isolated words with the help of acoustic model. The system was trained for 30 Hindi words. Training data had been collected from eight speakers. The experimental results of the developed system showed the overall accuracy 94.63%.

Dilip Kumar et al (2014) [22] have built a speech recognition system for Hindi language and search Hindi text in web search engine and also the text can be search from the database server using application server. In their paper they have covered two concepts one is the conversion of Hindi voice into text and second searching the same text into web application. They used Hidden Markov Model (HMMs) and HTK toolkit for recognizing the voice and identify Hindi language. The system recognizes the isolated words using acoustic word model. The system was trained for many Hindi words. Training data used for the system had been collected from nine speakers. The experimental results showed that the overall accuracy of the system with few states in HMM Such a design makes it truly practical to use text conversion and it's searching over the internet.

N. Mishra et al (2011) [23] built a speaker-independent connected Hindi digits recognition system. In this paper they have presented the work related to the system. For that authors have preferred clean and noisy both the environments. For speaker-independent connected Hindi digits recognition in clean and noisy environments robust features such as Revised Perceptual Linear Prediction (RPLP), Bark frequency Cepstral coefficients (BFCC) and Mel frequency perceptual linear prediction (MF-PLP) are used. Authors compared the recognition performance of these features were compared with recognition performance of Mel frequency Cepstral coefficient (MFCC), ΔMFCC and Perceptual linear prediction (PLP) features. Among all other methods MF-PLP features have shown best recognition efficiency for both clean as well as for noisy database. MFCC features were calculated by using feature extraction tool of Hidden Markov model Toolkit (HTK) and all other features were calculated using Matlab and those were saved in HTK format. Also the Training and testing was done using HTK.

Sharmila et al (2012) [24] in their paper describes the making of a speech recognition system for Hindi language recognition with Hidden Markov Model Toolkit (HTK). HTK recognizes the isolated digits using acoustic digits model. They have trained the system with 10 Hindi digits. The data used for Training data was collected from twenty four speakers. The system gives good accuracy in the range 93 - 100%.

3.4 Assamese Language

Himanshu sarma et al (2014) [25] described the Development of Assamese Speech Corpus and Automatic Transcription Using HTK. In this paper they have reported automatic transcription of Assamese speech using HTK. They have developed their database with the help of total 27 speakers. 14 males and 13 females within an age group 20 to 40 years. For performing feature extraction MFCC is used. Authors have transcribed recorded speech files using International Phonetic Alphabet (IPA) symbols and also in American Standard Code for Information Interchange (ASCII) for automatic transcription. They have used 527 and 127 files respectively for training and testing purpose. For that they have received accuracy of 65.65% with 38 phones.

3.5 Bengali Language

Biswajit Das et al (2011) [26] presents the Bengali Speech Corpus for Continuous Automatic Speech Recognition System. For this they have developed their own database which was divided in two age groups. First is younger group it contains 20 to 40 years age individuals and second older group it contains 60 to 80 years age individuals. They have used HTK and CMU-SPHINX speech recognition toolkit for training and testing of speech recognition system. For feature extraction purpose they have used MFCC procedure. They conclude that the ASR performance is depends on the quality of speech corpus.

3.6 Kannada Language

Shashidhara Nimbargi et al (2015) [27] in their paper have discussed about the ASR system which they had developed for Kannada language. They have used Hidden Markov model (HMM) and Mel Frequency Cepstral Coefficients (MFCC). For implementation HTK Toolkit was used. Also they have explained briefly various concepts like Speech Recognition Systems, HMM, Hidden Markov Model Toolkit (HTK), Speech data collection etc. the proposed system was able to identify the speech from various speakers and gave the accuracy between 83 to 100% for trained words.

G. Hemakumar et al (2013) [28] in their paper discussed about the designed algorithm which recognizes spoken Kannada words independent of speakers. The method which authors had used normalizes the original speech signal of every isolated word and extracts Linear-Predictive coding (LPC) coefficients, and also it converts them into Real Cepstrum Coefficient. After that these Real Cepstrum Coefficient values are subjected to dimensionality reduction through normal fit. These coefficients were used as the representatives of each spoken word. Euclidian distance measure was then used to compute the distance between the test samples to the model data in the database. The model datum in the database at a minimum distance is declared as the recognized word. For experimentation, they have used 294 unique Kannada words. Each of these words was recorded with 10 Speakers yielding 2,940 samples in total. Out of 10 speakers' data, 8 speakers' data i.e., 2,352 samples

were used to compute the representative co-efficient for each word. Remaining 2 speakers' data along with re-recorded data of two speakers out of the8 speakers is used for testing. Totally 2,352 signals are used for training and 1,176signals are used for testing. The success rate of the proposed system-known speaker data is 98.29 % and unknown speaker data is 91.66 %.

3.7 Malayalam Language

Smrithy K. Mukundan (2014) [29] in her paper had discussed the development of an Isolated Speaker Independent Automatic Speech Recognition system (ASR) for Malayalam language. Authors have used the Hidden Markov Model Toolkit (HTK) for implementing the system. Also they have used Hidden Markov Model (HMM) for acoustic modeling and Mel-Frequency Cepstral Coefficient (MFCC) as feature extraction. The system was trained with 21 speakers (8 male, 8 female and 5 children) who belongs to the age group from 4 to 76 years. The database used contains 210 isolated spoken words. They have taken separate utterance of each Malayalam words for the numbers 0 ('poojyam') to 9 ('onpathu'). For the purpose of training and testing of the system, author has divided the database into three equal parts. The experiment was conducted for both speaker dependent and speaker independent mode. The overall accuracy showed by the system for speaker independent mode ranges from 84% to 88% and for speaker dependent mode 94% to 100%.

Maya Moneykumar et al (2014) [30] presented the work about a syllable based word identification system for Malayalam using HTK. They have performed the task on the Linux platform. They have used HMM and MFCC techniques. The implementation of the system has been done using Hidden Markov Model Toolkit (HTK). They have trained the system using 40 vocabularies of bi-syllable words. They have trained the system with utterances of 3 male and 2 female speakers. The database used contains 40 utterances. For testing they have divided test data into two equal groups namely A and B which consists of 25 words each. While testing they have divided test data into two equal groups namely A and B which consists of 25 words each. So the system showed an accuracy of 76% and 80% respectively for the groups A and B.

3.8 Manipuri Language

Rahul. L. et al (2013) [31] have discussed in their paper about implementation of phoneme. They have used HMM tool kit (HTK), version 3.4 for implementation the system. A five state Hidden Markov Model (HMM) left to right with 32 mixture continuous density diagonal covariance Gaussian Mixture Model (GMM) per state was used to build a model for each phonetic unit. For developing the database data of around 5 hr read was collected from 4 male and 6 female speakers. Also for analyzing the systems performance Continuous Speech data it was collected from 5 males and 8 females. Total 69 words were chosen for the database. Using those chosen words sentences were framed for the purpose of recognizing those keywords by the system. For transcription of data the symbols of International Phonetic Alphabet (IPA) (which was revised in 2005) were used. An overall performance the system has shown after analysis was 65.24%

3.9 Punjabi Language

Kumar Ravinder (2010) [32] has worked for development speaker-dependent, real-time, isolated word recognizer for Punjabi language. Further he has extended his work up to comparison of speech recognition system for small vocabulary of speaker dependent isolated spoken words Punjabi language. He has used the Hidden Markov Model (HMM) and Dynamic Time Warp (DTW) techniques. As Punjabi language gave us the changes between consecutive phonemes detection of end point becomes highly difficult. The work presented gives the focus on template-based recognizer approach by using linear predictive coding with dynamic programming computation and vector quantization. With that it uses Hidden Markov Model based recognizers in isolated word recognition tasks, and it also reduces the computational costs. The parametric variation in the feature vector gave enhancement for recognition of 500-isolated word vocabulary on Punjabi language, as the Hidden Marko Model and Dynamic Time Warp technique gives 91.3% and 94.0% accuracy respectively.

Mohit Dua et al (2012) [33] worked for the implementation of an Automatic Speech Recognition system (ASR) for isolated word for an Indian regional language Punjabi. For the implementation the HTK toolkit based on Hidden Markov Model (HMM), a statistical approach, was used. At the beginning the system was trained for 115 distinct Punjabi words. For the database data was collected from eight speakers and for testing the system samples collected from six speakers in real time environments was used. Authors have developed a GUI for the implementation of the testing module using JAVA platform for making the system more interactive and fast. Also the authors have described the use of HTK Tool in different stages of development of system by presented a detailed architecture of an ASR system which was developed using HTK library modules and tools. After the testing phase the proposed system has showed the overall system performance was 95.63% and 94.08%.

Divya Bansal et al (2012) [34] in their paper described a Hidden Markov Model-based Punjabi text-to-speech (TTS) synthesis system (HTS). This system Hidden Markov Models generate speech waveform and applies it to Punjabi speech synthesis with the help of general speech synthesis architecture of HTK Tool Kit. Then this TTS system which I based on HMM can be used in mobile phones for storing phone directory or messages. Text messages and caller's identity in English language are mapped to tokens in Punjabi language. For building the synthesizer they have recorded the speech database and then phonetically segmented it. So first by extracting context-independent monophones and then context-dependent triphones. These speech utterances and their phone level transcriptions i.e.

monophones and triphones are the inputs to the speech synthesis system. System outputs the sequence of phonemes after resolving various ambiguities regarding selection of phonemes using word network files. For training the system speech corpora containing utterances of 61 words by one female speaker. And training data of 81 words.

3.10 Tamil Language

Ganesh A. A. et al (2013) [35] have presented their work about a syllable based speech recognition system for Tamil language. Tamil is a syllable based language. For the model described in this paper input was categorized into two categories first one was connected word which were segmented into individual Words using short term energy and second the isolated words which were further broken down into characters using Varied-Length Maximum Likelihood (VLML) algorithm. They have trained the system with 20 words which is related to agriculture domain and uttered by 2 speakers 2 times. They have used Gaussian Mixture Model (GMM), which is a speaker-independent model suitable for large sets of data, for classifying the characters for later pattern matching against the trained syllables. Also authors have introduced a new algorithm named VLML algorithm for identifying the boundary of each character. Accuracy given by the particular system was 70%.

K. Murali Krishna et al (2014) [36] They had proposed speech feature vector which was generated by projecting an observed vector onto an Integrated Phoneme Subspace (IPS) which was further based on Independent Component Analysis (ICA) or Principal Component Analysis (PCA). For Isolated Tamil Word Speech Recognition the performance of the new feature has to be evaluated. The proposed method was expected to provide higher recognition accuracy than conventional method in clean environment. They have used HTK for building the system. The average performance of the system using MFCC feature was in the range of 87% to 88% with word error rate 12% to 13%.

3.11 Telugu Language

P. VijaiBhaskar et al (2012) [37] worked for building a speech recognition system for Telugu language. Authors have used Hidden Markov Model Toolkit (HTK) to develop the system. The system was trained for continuous Telugu speech, the continuous Telugu speech data was recorded by male speakers. MFCC was used for generating the MFCC coefficient which is used for characterizing various speech sounds.

Surabhi Sreekanth et al (2005) [38] in their paper presented work about text-dependent speaker recognition system for Telugu which was designed by keeping in mind a low security access control systems. An isolated word speech recognition system was used to recognize the spoken password and then a speaker identification system was used to confirm the identity of the user from a known set of users. They have used HTK tool kit for building these systems. Also Hidden Markov Models based on Mel Frequency Cepstral Coefficients had been used to build models. Mahalanobis distance measure is employed. Authors used data of 7 speakers with 20 utterances. The system was trained by 10 words and tested by 10 words. For testing the testing was carried out in two cases first, for the isolated-word speech recognition and second for the second speaker recognition. For speech recognition the system was tested for password recognition. The system was built with speech samples of all the speakers with different passwords and the accuracy measured as, for 70 test samples it was 98.57%. And in second case i.e. speaker recognition for Different passwords for different speakers, It was observed that the average performance of the system was 93.71% and same password for all speakers, It was91.85%.

3.12 Gujarati Language

Jinal H. Tailor et. al. (2016) [39] in their paper presented architecture of ASR for Gujarati language. The database used for training purpose was collected from 4 male and 2 female who belongs to age group between 18 to 36 years. For measuring the performance and error parameters the authors have used Hidden Markov Model Toolkit. The system implemented analyzes WR (Word Recognition Rate) 95.9% and WER (Word Error Rate) as 5.85 % in Lab environment and in the open noisy environment calculated WR was 95.1% and WER found 7.40%.

3.13 English Language

The paper of Ahmad A. M. Abushariah et al (2010) [40] presents the design and implementation of English digits speech recognition system using Matlab (GUI). The presented system is based on the Hidden Markov Model (HMM). This system can recognize the speech waveform with the translating the speech waveform into a set of feature vectors using Mel Frequency Cepstral Coefficients (MFCC) technique. In this paper the focus is given on all English digits from (Zero through Nine), which are based on structure of isolated words. They have divided the work into two modules first one was the isolated words speech recognition and second is the continuous speech recognition. They had tested this system in clean as well as noisy environment and they observed that the system was showing a successful recognition rates those are In clean environment for isolated words speech recognition module, the multi-speaker mode have achieved 99.5% and the speaker independent mode have achieved 79.5%. Whereas in noisy environment both the modes have achieved 88% and 67% accuracy respectively. Also for continuous speech recognition module in clean environment, the multi-speaker mode have achieved 72.5% and the speaker-independent mode have achieved 56.25%.whereas in noisy environment both the modes have achieved 82.5% and 76.67% accuracy respectively.

Pawar et al. (2014) [41] has developed a speech recognition system for isolated digits of English language using HTK. The system developed by them is useful to both developers as well as researchers who are doing research in

English language for English digits or English words for the purpose of speech recognition. The authors have used two different corpora for this system. In both the audio recordings the speaker speaks isolated English language digits. Both corpora contains the training data and testing data. One corpus is self-recorded signals and other is standard Clemson University Audio Visual Experiments (CUAVE) dataset (50 speakers, each uttered 10 words). They have used the HTK toolkit for implementing the system. They have trained HMMs of the words by making the vocabulary on the training data. After testing the trained system on training data and test data the results evaluated shows that 95% of the accuracy was achieved.

Pawar et al. (2015) [42] have presented work related to recognition of isolated English digits. The system is based on Mathematical tool like HTK. They have used two databases which contains the recordings of isolated English Digits from ZERO to NINE. They had implemented the system for speaker dependent and speaker independent method and their comparison has been done. They had used a separate Hmm for every digit. After training the system had been checked for accuracy and it had given successful results also. And the system can also be used for recognition of continuous English words.

Table 1.1 Comparison of various languages

Sr. No.	Language	Authors	Year	Type of utterance	No. of Speakers	Utterances of each word	Recording environment	No. of words / sentences	Feature extraction technique used	Word accuracy	No. of states in HMM	Word error rate
1	SANSKRIT	Jitendra Singh Pokhariya et al.	2014	Isolated	10 speakers	1	-	198 words	MFCC	95.2 % 97.2 %	5 10	4.8 % 2.8 %
2.	AHIRANI	Ajay S. Patil	2014	Isolated	20 speakers	5	Room Environment	20 words	MFCC	94%	6	6%
3	HINDI	Shweta Tripathy et al	2013	Isolated	5 (2 male 3 female)	7	Room Environment	35 Hindi words	MFCC LPC	76.44% 27.02%	5	23.56% 72.98%
		Annu Choudhary et al	2013	Connected	1	20	Room Environment	20 Connected word	MFCC	90.00%	5	10.00%
				Isolated	1	10	Room Environment	100 distinct Hindi words	MFCC	95%	4	5%
		Ankit kumar et al	2014	Continuous	1	10	Room Environment	70 distinct Hindi words	MFCC	95.08%	5	5%
		Gaurav et al.	2012	Continuous	12 females and 18 males	1	Office Noise Condition	43 distinct sentences	MFCC	88.81	5	11.19
		Preetisaini et al	2013	Isolated	9 (5 males and 4 females)	3	Room Environment	116 Hindi words	MFCC	95 to 96 %	10	6 to 8 %

		Babita Saxena et al	2015	Isolated	10	Not Available	Varying Noise Environments	10 Hindi Digits	MFCC	94.09%	5	5.91%
		Kuldeep Kumar et al	2011	Isolated	8(5 male and 3 female)	4 times	Room Environment	30 words	MFCC	94.63%	5-11 state	5.37%
		Dilip Kumar et al	2014	Isolated	5 males and 4 females	3	Specified Environment	113 words	MFCC	-	-	-
		Sharmila et al	2012	Isolated	24 speakers	1	Room Environment	10 Hindi Digits	MFCC	93-100 %	3-5	0-7%
4	ASSAMESE	Himansu sarma et al	2014	Continuous	27 (14 male and 13 female)	1	In A Noise Free Room	Not Specified as context is different	MFCC	65.26%	-	34.74%
5	BENGALI	Biswajit Das et al	2011	Continuous	70 male and 40 female	-	-	7500 unique sentence i.e 19640 unique words	MFCC	-	5 states	-
6	KANNADA	Shashidhara Nimbargi et al	2015	Isolated	35 (19 male and 16 female)	1	Controlled Noise Condition Environment	31 words	MFCC	83 – 100 %	-	17 – 0%
		G. Hemakumar et al	2013	Isolated	10 speakers	-	Room Environment	294 Kannada words	LPC	known Speaker data is 98.29 % Unknown speaker data is 91.66 %.	-	1.71 % 8.34 %

7	MALAYALAM	Smrithy K Mukundan	2014	Isolated	21 speakers (8 male, 8 female and 5 children)	1	Normal Environment	10 digits	MFCC	for speaker independent mode ranges from 84% to 88% for speaker dependent mode ranges from 94% to 100%.	5 states	for speaker independent mode ranges from 16% to 12% For speaker dependent mode ranges from 6% to 0%
		Maya Moneykumar et al	2014	Isolated	3 male and 2 female	1	Noisy Environment	40	MFCC	76- 80 %	5	24 – 20%
8	MANIPURI	Rahul. L. et al	2013	Isolated	13 (8 male and 5female)	1	Noiseless Environment.	69 words	MFCC	65.24%	5 states	34.76 %
9	PUNJABI	Kumar Ravinder	2010	Isolated	-	10	-	500 words	LPC	91.3 %	-	8.7%
		Mohit Dua et al	2012	Isolated	8	3 times	Real Time Environment	115	MFCC	94 % to 96%	4 states	4% to 6%.
		Divya Bansal	2012	Isolated	1 Female	1 Utterance	Room Environment	61Words 81 Words	MFCC	71% 80%	5 States	29% 20%
10	TAMIL	Ganesh A. A. et al	2013	Isolated	Speakers (2 female and 2 male).	5 times	Noiseless Room Environment	20 Words	MFCC	70%	-	30%

		K.Murali Krishna et al	2014	Isolated	3 speakers	9 times	In A Normal Room With Minimal External Noise	5 words	MFCC	87% to 88%	-	12% to 13%.
11	TELUGU	P. VijaiBhaskar et al	2012	continuous	-	-	-	29 context-dependent Telugu phonemes	MFCC	-	5 states	-
		Surabh Sreekanth et al	2005	Isolated	7	20 times	-	1 word	MFCC	91.85 - 93.71 %	-	8.15 – 6.29 %
12	GUJARATI	Jinal H. Tailor et al	2016	Isolated	6 (4 male and 2 female)	-	noisy environment	10-12 words	MFCC	95.9%	-	5.85 %
							quite laboratory environment			95.1%		7.40%
13	ENGLISH	Ahmad A. M. Abushariah et al	2010	Isolated	34 (24 male and 10 female)	-	In Both Noise Free (Clean)	10 distinct words (digits)	MFCC	Multi-Speaker Mode 99.5 AND Speaker-Independent Mode 79.5	-	Multi-Speaker Mode 0.5 AND Speaker-Independent Mode 20.5
							Noisy Environment			Multi-Speaker Mode 88 AND Speaker-		Multi-Speaker Mode 12 AND Speaker-

										Independent Mode 67		Independent Mode 33
				Continuous			In Both Noise Free (Clean)			Multi-Speaker Mode 92.5AND Speaker-Independent Mode 76.67		Multi-Speaker Mode 7.5 AND Speaker-Independent Mode 23.35
							Noisy Environment			Multi-Speaker Mode 72.5 AND Speaker-Independent Mode 56.25		Multi-Speaker Mode 27.5 AND Speaker-Independent Mode 43.75
		Pawar et al.	2014	Isolated	50	1 times	-	10	MFCC	95%	7	5%
		Pawar et al.	2015	Isolated	36	-	-	10 digits	MFCC	96% for speaker dependent	-	4% for speaker dependent
										86% for speaker independent		14% for speaker independent

4. CONCLUSIONS

In this paper we have discussed some of the papers which are related to Automatic speech recognition systems which are built using HTK toolkit for different Indian regional Languages. we have also compared all those systems on the basis of their Language, Year, Type of utterance, No. of Speakers, Utterances of each word, Recording environment, No. of words / sentences, Feature extraction technique used, Word accuracy, No. of states in HMM, Word error rate. For Indian languages the work done is less as compared to other foreign languages. As we have seen the systems which are built using HTK toolkit have high accuracy rate as compared to other techniques. This study will motivate people for developing ASR systems using HTK toolkit for different languages.

ACKNOWLEDGEMENT

This work is supported by University Grants Commission under the scheme Major Research Project entitled as "Development of Database and Automatic Recognition System for Continuous Marathi Spoken Language for agriculture purpose in Marathwada Region". The authors would also like to thank the Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad for providing the infrastructure to carry out the research.

REFERENCES

[1] Devyani S. Kulkarni, Ratnadeep R. Deshmukh, and Pukhraj P. Shrishrimal. "A Review of Speech Signal Enhancement Techniques." International Journal of Computer Applications 139.14 (2016).

[2] V. B. Waghmare, R. R. Deshmukh, P. P. Shrishrimal and G. B. Janvale "Development of Isolated Marathi Words Emotional Speech Database" International Journal of Computer Applications 94(4):19-22, May 2014.

[3] Pukhraj P. Shrishrimal, Ratnadeep R. Deshmukh and Vishal B. Waghmare "Indian Language Speech Database: A Review" .International Journal of Computer Applications 47(5):17-21, June 2012.

[4] Kandagal, Amaresh P., and V. Udayashankara. "Automatic bimodal audio visual speech recognition: A review." Contemporary Computing and Informatics (IC3I), 2014 International Conference on. IEEE, 2014.

[5] Muskan, Naveen Aggarwal, "Punjabi Speech Recognition: A Survey", Proc. of the Intl. Conf. on Advances in Engineering and Technology - ICAET-2014.

[6] Jadhav A.V., Pawar R.V., "Review of various approaches towards speech recognition," in Biomedical Engineering (ICoBE), 2012 International Conference on , vol., no., pp.99-103, 27-28 Feb. 2012.

[7] Jianliang Meng, Junwei Zhang, Haoquan Zhao, "Overview of the Speech Recognition Technology," in Computational and Information Sciences (ICCIS), 2012 Fourth International Conference on , vol., no., pp.199-202, 17-19 Aug. 2012

[8] Ahmad A. M. Abushariah, Teddy S. Gunawan, Othman O. Khalifa, Mohammad A. M. Abushariah, "English Digits Speech Recognition System Based on Hidden Markov Models", International Conference on Computer and Communication Engineering (ICCCE 2010), Kuala Lumpur, Malaysia, 11-13 May 2010.

[9] M. A. Anusuya, S. K. Katti, "Speech Recognition by Machine: A Review", International Journal of Computer Science and Information Security (IJCSIS), Vol. 6, No. 3, pp. 181-205, 2009.

[10] Shweta Tripathy, Neha Baranwal, G. C. Nandi, "A MFCC based Hindi Speech Recognition Technique using HTK Toolkit", Proceedings of the 2013 IEEE Second International Conference on Image Information Processing ,(ICIIP-2013), 2013.

[11] R. L. Rabinner, B. H. Huang, "An Introduction To Hidden Markov Models," IEEE Acoust, Speech Signal Processing Mag., pp. 4-16, 1986.

[12] Kuldeep Kumar, R. K. Aggarwal, "HINDI SPEECH RECOGNITION SYSTEM USING HTK", International Journal of Computing and Business Research ISSN Online): 2229-6166 Volume 2 Issue 2 May 2011.

[13] http://spandh.dcs.shef.ac.uk/ed_arena/htk/index.html.

[14] Jitendra Singh Pokhariya, D r. Sanjay Mathu r, "Sanskrit Speech Recognition using Hidden Markov Model Toolkit", International Journal of Engineering Research & Technology (IJERT) Vol. 3 Issue 10, October- 2014.

[15] Ajay S. Patil, "Automatic Speech Recognition for Ahirani Language Using Hidden Markov Model Toolkit (HTK)", International Journal of Computer Science Trends and Technology (IJCST) – Volume 2 Issue 3, May-Jun 2014.

[16] Annu Choudhary, Mr. R.S. Chauhan, Mr. Gautam Guptam, "Automatic Speech Recognition System for Isolated & Connected Words of Hindi Language By Using Hidden Markov Model Toolkit (HTK)", Proc. of Int. Conf. on Emerging Trends in Engineering and Technology, organized by ACEEE, 2013.

[17] Ankit Kumar, Mohit Dua, Tripti Choudhary,"Continuous Hindi Speech Recognition Using Monophone based Acoustic Modeling", International Journal of Computer Applications® (IJCA) (0975 – 8887) International Conference on Advances in Computer Engineering & Applications (ICACEA-2014) at IMSEC,GZB.

[18] Gaurav, Devanesamoni Shakina Deiv, Gopal Krishna Sharma, Mahua Bhattacharya, "Development of Application Specific Continuous Speech Recognition System in Hindi", Journal of Signal and Information Processing, 2012, 3, 394-401

[19] Preeti Saini, Parneet Kaur, Mohit Dua, "Hindi Automatic Speech Recognition Using HTK", International Journal of

Engineering Trends and Technology (IJETT) - Volume4 Issue6- June 2013.

[20] Babita Saxena, Charu Wahi,"Hindi Digits Recognition System on Speech Data Collected In Different Natural Noise Environments", International Conference on Computer Science, Engineering and Information Technology (CSITY 2015) February 14~15, 2015, Bangalore, India. Volume Editors: David C. Wyld, Jan Zizka ISBN : 978-1-921987-31-1

[21] Kuldeep Kumar, R. K. Aggarwal, "Hindi Speech Recognition System Using HTK", International Journal of Computing and Business Research, Volume 2 Issue 2 May 2011.

[22] Dilip Kumar, Abhishek Sachan, Malay Kumar, "Implementation of Speech Recognition in Web Application for Sub Continental Language", International Journal of Engineering Trends and Technology (IJETT) – Volume 9 Number 11 - Mar 2014.

[23] A. N. Mishra, Mahesh Chandra, Astik Biswas, S. N. Sharan, "Robust Features for Connected Hindi Digits Recognition", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 4, No. 2, June, 2011.

[24] Sharmila, Dr. Neeta Awasthy, Dr. R. K. Singh, "Performance of Hindi Speech Isolated Digits In HTK Environment", IOSR Journal of Engineering May. 2012, Vol. 2(5) pp: 1020-1023

[25] Sarma, Himangshu, Navanath Saharia, and Utpal Sharma. "Development of Assamese Speech Corpus and Automatic Transcription Using HTK." Advances in Signal Processing and Intelligent Recognition Systems. Springer International Publishing, 2014. 119-132.

[26] Biswajit Das, Sandipan Mandal, Pabitra Mitra, "Bengali Speech Corpus for Continuous Automatic Speech Recognition System" 978-1-4577-0931-9/11/$26.00, 2011 IEEE.

[27] Shashidhara Nimbargi, Dr. S. N. Chandrashekara, "Isolated Speaker Independent Kannada ASR System Using HTK", International Journal of Combined Research & Development (IJCRD) eISSN: 2321-225X;pISSN: 2321-2241 Volume: 4; Issue: 6; June -2015.

[28] G. Hemakumar, P. Punitha, "Speaker Independent Isolated Kannada Word Recognizer", Multimedia Processing, Communication and Computing Applications Volume 213 of the series Lecture Notes in Electrical Engineering pp 333-345, Date: 26 May 2013.

[29] Smrithy K Mukundan, " 'ShreshtaBhasha' Malayalam Speech Recognition using HTK", International Journal of Advanced Computing and Communication Systems (IJACCS) vol.1 Issue.1 March 2014. ISSN: 2347 – 9299 / 2347 – 9280.

[30] Maya Moneykumar, Elizabeth Sherly ,Punjabi, Marathi ,"Malayalam Word Identification For Speech Recognition System", An International Journal of Engineering Sciences, Special Issue iDravadian , December 2014, Vol. 15 ISSN: 2229-6913 (Print), ISSN: 2320-0332 (Online) -, Web Presence: http://www.ijoes.vidyapublications.com © 2014 Vidya Publications.

[31] Rahul, L.; Nandakishor, S.; Singh, L.J.; Dutta, S.K., "Design of Manipuri Keywords Spotting System using HMM," in Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013 Fourth National Conference on , vol., no., pp.1-3, 18-21 Dec. 2013.

[32] Kumar Ravinder, "Comparison of HMM and DTW for Isolated Word Recognition System of Punjabi Language", Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications ,Volume 6419 of the series Lecture Notes in Computer Science pp 244-252, 2010

[33] Mohit Dua, R. K. Aggarwal, Virender Kadyan, Shelza Dua, "Punjabi Automatic Speech Recognition Using HTK",IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012.

[34] Divya Bansal, Ankita Goel, Khushneet Jindal, "PUNJABI SPEECH SYNTHESIS SYSTEM USING HTK", International Journal of Information Sciences and Techniques (IJIST) Vol.2, No.4, July 2012.

[35] Ganesh, A.A.; Ravichandran, C., "Grapheme Gaussian model and prosodic syllable based Tamil speech recognition system," in Signal Processing and Communication (ICSC), 2013 International Conference on , vol., no., pp.401-406, 12-14 Dec. 2013.

[36] K.Murali Krishna, M.Vanitha Lakshmi, "Speaker Independent Isolated Tamil Words for Speech Recognition using MFCC, IPS and HMM", International Journal of Scientific & Engineering Research, Volume 5, Issue 4, April-2014.

[37] P. Vijai Bhaskar,S. Rama Mohan Rao, A. Gopi, "HTK Based Telugu Speech Recognition", International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 12, December 2012.

[38] Surabhi Sreekanth, Kavi Narayana Murthy, "Text-Dependent Speaker Recognition System for Telugu" Osmania Papers in Linguistics, Vol.31, 2005, 84-99.

[39] Jinal H. Tailor, Dipti B. Shah, "Speech Recognition System Architecture for Gujarati Language", International Journal of Computer Applications (0975 – 8887) Volume 138 – No.12, March 2016.

[40] Ahmad A. M. Abushariah, Teddy S. Gunawan, Othman O. Khalifa, Mohammad A. M. Abushariah, "English Digits Speech Recognition System Based on Hidden Markov Models", International Conference on Computer and Communication Engineering (ICCCE 2010), 11-13 May 2010, Kuala Lumpur, Malaysia.

[41] Ganesh S Pawar, Sunil S Morade, "Isolated English Language Digit Recognition Using Hidden Markov Model Toolkit", International Journal of Advanced Research in

Computer Science and Software Engineering 4(6),June - 2014, pp. 781-784.

[42] Ganesh S Pawar, Sunil S Morade, "Isolated Digit Recogntion using Mathematical Tool based on Hidden Markov Model", International Journal of Modern Trends in Engineering and Research (IJMTER) Volume 2, Issue 7, [July-2015] Special Issue of ICRTET'2015.