Development of Automatic Kannada Speech Recognition System

Akshata K Shinde[1,] Anjali H R[2], Deepika N Karanth[3], Gouthami K[4], Vijetha T S[5]

[1234]4th Year, Department of Electronics and Communication
[5]Assistant Professor, Department of Electronics and Communication, Alva's Institute of Engineering and Technology, Mijar

Abstract

Automatic Speech Recognition (ASR) is the technology that allows human being to use their voices to speak with a computer interface in a way that, in its most sophisticated variations, resembles normal human conversation. Automatic speech recognition is an area of research which deals with the recognition of speech by machine in several conditions. This paper presents a brief survey on Automatic Kannada Speech Recognition Systems. Many research and developments are made to improve the performance of ASR to work more efficient by the researchers. The criteria for designing Speech Recognition system are data preparation, pre-processing filter, end-point detection, feature extraction techniques, speech classifiers, and performance evaluation. Speech recognition system for Kannada language has been implemented using the Hidden Markov Tool Kit (HTK).The objective of this review paper is to summarize the researches made by researchers in the recent year.

***Keywords*:** *Automatic Speech Recognition System (ASR), Mel frequency cepstral coefficients (MFCC), Hidden Markov Model (HMM), Hidden Tool Kit (HTK).*

I. INTRODUCTION

Over the years, the development of speech recognition applications gives extremely useful contributions to this field of research and is becoming more mature in recent years. Speech is one of the fundamental way of communication mode of the human and, manmachine interaction is almost become basic necessity in this era. Speech processing is the study of speech signals and the different methods for processing of signals. The signals are normally processed in a digital representation, so speech processing can be considered as a special case of digital signal processing, applied to speech signals. Developing any kind of speech recognition system which can convert speakers spoken words in to particular form which a computer can understand and do further processing based on the application requirements can be a very useful tool. Automatic Speech Recognition (ASR) is the technology that allows human being to speak with a computer interface in a way that, in its most sophisticated variations, resembles normal human conversation.

Kannada is a language spoken in India predominantly in the state of Karnataka and is spoken by about 60 million speakers. This language is also spoken in neighboring states like Maharashtra, Tamil Nadu, Andra Pradesh, Goa etc. Karnataka has different dialect regions also. However, there is only little research reported on Kannada speech processing compared to other languages of similar importance.

The motive of this paper is to render background information on the factors to be considered here and also it stresses to check if the there is any resemblance in the work carried out by other authors. Today's general-purpose speech recognition systems are based on Hidden Markov Models (HMM). These are statistical models that gives a sequence of symbols or quantities. Speech recognition uses HMMs because a speech signal can be viewed as a piece wise stationary signal or a short-time stationary signal.

II. LITERATURE SURVEY

Kannadaguli and Thalengala [1]**,** presented a speaker dependent Hidden Markov modelling for phoneme recognition in Kannada language using stochastic pattern recognition and Acoustic phonetic schemes to recognize phonemes. They have used native Kannada language. They made use of Mel – Frequency Cepstral Coefficients for speech feature extraction. The performance analysis is done in terms of Phoneme Error Rate (PER). The template classifier and pattern recognizer are built using Hidden Markov models. End point detection, automatic silence removal of the training samples are performed with an algorithm based approach. During the recognition phase, the speech features of an unknown utterance are searched against the HMM phoneme models trained. Each of the valid combinations of Kannada phonemes that results in Kannada vowels or consonants can be successfully modelled using HMM.

Droppo and Acero [2], has stated that in the modern speech recognition systems, several hidden Marco models are divided into a physical cluster state, here different states cannot share the same cluster. Authors have conducted three experiments those are based on Aurora-1 Aurora 2 and Aurora 3 to overcome the above mention limitations. Decision trees are the common tools used in clustering. Global decision tree are used to cluster the logical states into acoustic models in all in three different experiments. The accuracy is increased by reducing the model size.

Dai and Soon [3], suggest an adaptive voice activity detected algorithm is implemented in a MFCC based speech recognition system. The system describes a novel scheme of detecting speech presence or absence by tracking the higher portion of speech power spectrum. Evaluation tests have been conducted based on the Standard English database Aurora 2. The key feature of the proposed system is the clean speech variance. The threshold is generated from the speech which leads to the ability to adapt to changing environments. The experiments which are conducted shows the significant improvement in the recognition accuracy.

Pradeep and Rao [4], have urge about Deep Neural Network (DNN) based speech recognizers have recently replaced Gaussian Mixture Model (GMM). Developing a phonetic engine and enhancing its performance can lead to significant improvement in Automatic speech recognition. It gives the comparative study of speech recognition baselines HMM-GMM, HMM-ANN and HMM-DNN are analyzed. First set of experiments use the Kannada speech corpus, which contains continuous utterances recorded in three different modes namely read mode, lecture mode and conversation mode. Context independent acoustic modeling is built on the conventional HMM-GMM and HMMANN baselines and evaluated on different modes of the Kannada speech corpus. The target labels to the HMM-ANN baseline is obtained by force aligning the Training speech corpus using HTK. The learning phase of DNN undergoes two stages; unsupervised pre-training followed by fine tuning stage. HMM-DNN baseline improves the PER by an amount of nearly 7-8 % as compared to the conventional HMM-GMM and HMM-ANN baselines.

Giurgiu and kabir [5], have proposed the ability of the Hidden Marko model based method to generate accurate automatic phone-level transcriptions for vocabulary speech corpus ROGRIP. The automatic time-aligned phone transcription toolbox has been tested on Romanian corpus and validated on the

English database. The quality of transcriptions is decided by evaluating the statistical parameters of the error between the automatic and manual transcription. The transcriptions obtained from the most reliable system deviate from the average manual transcription by an average of 20ms. The system is also able to convert the generated transcription from HTK format into PRAAT format**.** The automatic transcription performs well over the ROGRID data in the comparison to manual transcription. The limitation of the proposed system is it requires word level transcription as input and produces time aligned phone level transcription as the output.

M. V. Shridhara [6], presented on the system development and availability of spoken language corpora in regional languages. The issues of regional bias, accent, unique style and diversity associated with each geographical region and language will have a significant effect on the performance of speech recognition/synthesis systems. This model collection of speech data in Kannada language for prosodically guided phonetic search engine and the issues involved in transcription. The speech corpus contains data in three different modes namely, read mode, conversation mode and extempore mode. A baseline recognition system for Kannada language is designed and constructed using HTK for the data collected in different modes.

H. Muralikrishna [7], presented the method for implemented Kannada isolated digit recognition system using Mel frequency cepstral coefficients. The system is built to recognize isolated utterances of Kannada numbers. MFCC are used as the features and Hidden Markov Model as pattern recognizer. K-means procedure is performed on the feature vectors to get the observation sequence. Discrete HMM is used in the system. The system is constructed by considering the requirement of a voice controlled machine in Kannada language. Performance of the system is analyzed and compared based on the MFCC along with its first and second order derivatives.

Kannadaguli and Bhat [8]**,** presented the comparison between Gaussian Mixture Modelling (GMM) and Hidden Markov modeling (HMM) for phoneme recognition systems. Both models were built by using stochastic pattern recognition and Acoustic phonetic schemes to recognise phonemes. They have used Kannada language to train and test these models. They made use of Mel – Frequency Cepstral Coefficients. The performance analysis of both models is done in terms of Phoneme Error Rate (PER) justifies the fact that HMM yields better results over GMM. Template classifier and pattern recognizer are built using these two methods, which are of stochastic approaches. The results it can be inferred that, each of the valid combinations of
Kannada phonemes can be successfully modelled using HMMs, with one emitting state per model. Results reveal that, this method is suitable for building automatic phoneme recognition systems.

Shashidhara Nimbargi [9], stated that Kannada language is one of the oldest language of world. As per estimation there are more than 40 million people who speak Kannada, with this there is further research in Kannada speech recognition system for development of commercial as well as social purposes. According to review, MFCC has better recognition rate then LPCC techniques. The basic ASR system is used HTK. Therefore Kannada speech recognition system using HTK was implemented as it was very effective and simple. Technology of speech recognition have extended from speaker dependent to speaker independent system by LVCSR (Large Vocabulary Continuous Speech Recognition System).

Krishna et al [10], has suggested a speech recognition system for Kannada language that has been implemented using the Hidden Markov Tool Kit (HTK). The system performance is comparatively studied and evaluated for syllable and phone level models. The Kannada word dictionary of size about 110 words is used in the study and Mel frequency cepstral coefficients are computed in acoustic front-end processing. The system is designed to recognize isolated utterances of Kannada words. Baum-Welch algorithm is used to train the Hidden Markov model and Viterbi algorithm for decoding process. The

objective is to compare the performances of phone-level and syllable-level acoustical models for Kannada language vocabulary. Average word recognition accuracy of 97.1% for syllable-level modelling and 98.6% for phone-level modelling has been reported.

Singhal and Dubey [11], have presented an Automatic speech recognition system for the connected words. The implementation of the ASR is done by extending an isolated word recogniser for the speaker independent data. The approach used is Mel frequency cepstral coefficient. For the back end mapping of unknown utterances, Hidden Markov model and Dynamic time wrapping are used. The data base of English and Hindi language words are created and sentences for testing. The result is based on the comparison between the Hidden Markov model and the Dynamic Time Warping model. The efficiency of the system for dividing the speech into words has been checked.

S B Harisha [12], implemented a model for the small vocabulary Kannada speech recognition consisting of eighteen words. Segmentation of speech into word is accomplished using energy, magnitude and spectrum. The developed system is tested for withdrawal slip filling application. The proposed system is designed and simulated using a MATLAB. The developed system achieved the recognition accuracy of about 91.5%. This system acts as a basis for real time speech recognition products, where input will never be an isolated word.

III. METHODOLOGY

The implementation of speech recognition system for Kannada words has been carried out using Hidden Markov Toolkit in the Linux platform. The implementation involves mainly data preparation, data preprocessing, acoustic modelling using HMM is done by using HTK tool kit and finally evaluating the performance of the system by testing. Speech data is needed both for training and for testing. The system to be built, requires a set of kannada sentences and its transcription and also its dictionary, corresponding pronunciation and phones data and all of the sentences need to be recorded from scratch and to do this scripts are needed to prompt for each sentence. Each recorded speech is then used for MFCC feature extraction. In the case of the training data, the prompt scripts will be used in conjunction with a pronunciation dictionary to provide the initial phone level transcriptions needed to start the HMM training process and then performance of the recognizer has to be evaluated on the testing data set. The result analysis should be on the percentage of recognition accuracy.

IV. CONCLUSION

This paper has reviewed the Automatic Kannada speech recognition system and the methods used in recognition system and techniques used by the authors to enhance the recognition system. The proposed project of speech recognition system for Kannada words can be carried out using Hidden Markov Toolkit in the Linux platform which is one of the best technique. Speech data is needed both for training and for testing. The system to be built, requires all of the speech recorded from scratch and to do this scripts are needed to prompt for each sentence. In the case of the training data, the prompt scripts will be used in conjunction with a pronunciation dictionary to provide the initial phone level transcriptions needed to start the HMM training process and testing should be done to evaluate the system accuracy.

V. REFERENCES

[1] Prashanth Kannadaguli, Ananthakrishna Thalengana, “Phoneme Modeling for Speech Recognition in Kannada Using Hidden Markov Model”, 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)978-1-4799-1823-2/15.

[2] Jasha Droppo and Alex Acero, "Context Dependent String Edit Distance for Automatic Speech Recognition", 2010 IEEE International Conferenceon Acoustics, Speech and Signal Processing,4358978-1-4244-4296-6/10.

[3] Peng Dai, Ing Yann Soon, "An Adaptive Soft Voice Activity Detector for Automatic Speech Recognition System", 2011 8th International Conference on Information, Communications & Signal Processing,978-1-4577-0031-6/11.

[4] Pradeep R, K. Sreenivasa Rao "Deep Neural Networks for Kannada Phoneme Recognitation" 978-1-5090-3251-8/16 2016 IEEE.

[5] Mircea Giurgiu and Ahsanul Kabir, "Automatic Transcription and Speech Recognition of Romanian Corpus RO-GRID", 2012 35th International Conferenceon Telecommunications and Signal Processing (TSP).

[6] M. V. Shridhara ""Development of Kannada speech corpus for prosodically guided phonetic search engine," 2013 International Conference Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (OCOCOSDA/CASLRE), Gurgaon, 2013, pp.1-6.doi: 10.1109/ICSDA.2013.6709875.

[7] H. Muralikrishna, "HMM based isolated Kannada digit recognition system using MFCC," 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Mysore, 2013, pp. 730-733.

[8] Prashanth Kannadaguli, Vidya Bhat, "A Comparison of Gaussian Mixture Modeling (GMM) and Hidden Markov Modeling (HMM) based approaches for Automatic Phoneme Recognition in Kannada", 2015 International Conference on Signal Processing and Communication (ICSC) 978-1-4799-6761-2/15.

[9] Shashidhar Nimbargi, "Isolated Speaker Independent Kannada ASR System using HTK" International journel of combined research and development (IJCRD) eISSN:2321225X;pISSN; 2321-2241 volume:4; issue: 6; June-2015.

[10] Ananthakrishna T, Maithri M, Dr. Kumara shama, " Kannada Word Recognition Sytem using HTK" , IEEE INDICON 2015 1570157619.

[11] Shweta Singhal Dr. Rajesh Kumar Dubey, "Surface Acoustic Wave Devices for Harsh Envirnment Wireless Sensing", 2015 Communication, Control and Intelligent Systems (CCIS),978-1-4673-7541-2/15.

[12] S B Harisha, "Kannada Speech Recognition Using MFCC and KNN Classifier for Banking Applications" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 1, January 2017.