Recent Trends in Speech Recognition approaches and its application for Telugu Language for Speech & Hearing Impaired

P. Jeethendra
Research Scholar
RAYALASEEMA UNIVERSITY
Kurnool - 518007 – AP (India)

M. Chandrashekar
Centre for Advanced Systems, (BM), Bharat Dynamics Limited,
Talangana, Hyderabad - India

Abstract :-In this paper, we compare the advantages and disadvantages of the major speech to text conversion techniques in both spectral and temporal domain. The aim of this paper is to analyze and provide a simple and precise view of the Speech to text conversions based on HMM, ANM. etc which can be adapted for the Speech to Text conversion exclusively for Telugu language. Recognizing and understanding the Speech is an important aspect for Speech and Hearing Impaired. For a Speech and Hearing Impaired person, it is difficult to comprehend the real time speech through listening only. A real time speech to text conversion system converts the spoken words into text form exactly in the similar way that the user pronounces. Conversion of speech to text helps to understand the speech in real time by using this visual reinforcement cue. For converting the speech to text there are various Automatic Speech Recognition (ASR) and speech enhancement techniques technological devices. The Assistive Technology involving voice / speech communication is used primarily by people who are deaf , or who have speech and / or Language impairment.

Keywords : Speech to text conversion, Speech enhancement techniques, accuracy of conversion, pattern matching

1. INTRODUCTION

Speech to Text research on local language to help the Speech and Hearing impaired peoples. Research in Telugu Speech recognition field is still in the nascent stage. A large amount of continuous work is needed in the field of Telugu Speech to text conversion. This paper is on the different techniques available for the speech to text conversion which can be adapted for the Telugu speech to text conversion systems, and a initiative has been taken to develop the primary understanding about the Telugu speech to text conversion and a brief information is provided based on literature survey.

Various researchers have discussed different approaches with slight modifications in the concept and algorithms. Mostly, it is observed that the core segment remains the same. Most of the work has been carried out in English language and to an extent in Hindi Language also has been done at C-DAC and IIIT Gwalior India. This paper initiates for a plan to develop the speech recognition in Telugu language, such that it will enhance the learning ability of the speech and hearing impaired, where learning in local language is advantageous.

A fundamental unit of any spoken language is the Phoneme(1), It is produced by a human vocal tract. These are elementary sounds of a language. The phoneme is distinctive in the sense that is speech sound class that differentiates words of languages. Speech scientists used the word "phone" to mean a particular instantiation of phoneme. The speech sounds are studied in the following two types. First, an articulatory phonetics and second, an acoustical phonetics

II. CONCEPT OF INDIAN LANGUAGE AND SCRIPT

The scripts of Indian Languages have about 35 Consonants and 18 Vowels in most of the Indian Languages. In consonants speech sound, the breath is at least partly obstructed and which can be combined with a vowel to form a syllable. Phonetic Nature is an important feature of Indian Languages scripts.. All Indian Language scripts have common phonetic base. The Indian languages can be mapped for sounds in a straight forward way. The table shows some common types and similarity in Indian phonetic languages

Vowels	DEV	GUJ	PUN	BEN	ORI	TEL	KAN	TAM	MAL	SIN	URD	SIND
ka	क	ક	ਕ	ক	କ	క	ಕ	க	ക	ක	ک	
kā	का	કા	ਕਾ	কা	କା	కా	ಕಾ	கா	കാ	කා	کا	
kæ	–	–	–	–	–	–	–	–	–	කැ	–	
kǣ	–	–	–	–	–	–	–	–	–	කෑ	–	
ki	कि	કિ	ਕਿ	কি	କି	కి	ಕಿ	கி	കി	කි	کِ	
kī	की	કી	ਕੀ	কী	କୀ	కీ	ಕೀ	கீ	കീ	කී	کی	
ku	कु	કુ	ਕੁ	কু	କୁ	కు	ಕು	கு	കു	කු	کُ	
kū	कू	કૂ	ਕੂ	কূ	କୂ	కూ	ಕೂ	கூ	കൂ	කූ	کُو	
kṛ	कृ	કૃ	–	কৃ	କୃ	కృ	ಕೃ	–	കൃ	කෘ	–	
kṝ	कॄ	કૄ	–	কৄ	କୄ	కౄ	ಕೄ	–	കൄ	කෲ	–	
ke(s)	के *	–	–	–	–	కె	ಕೆ	கெ	കെ	කෙ	–	
ke/ē	के	કે	ਕੇ	কে	କେ	కే	ಕೇ	கே	കേ	කේ	کے	
kai	कै	કૈ	ਕੈ	কৈ	କୈ	కై	ಕೈ	கை	കൈ	කෛ	کَے	
ko(s)	को *	–	–	–	–	కొ	ಕೊ	கொ	കൊ	කො	–	
ko/ō	को	કો	ਕੋ	কো	କୋ	కో	ಕೋ	கோ	കോ	කෝ	کو	
kau	कौ	કૌ	ਕੌ	কৌ	କୌ	కౌ	ಕೌ	கௌ	കൌ	කෞ	کَو	
kaḥ	कः	કઃ	ਕਃ	কঃ	କଃ	కః	ಕಃ	கஃ	കഃ	කඃ	–	–
kan	कं	કં	ਕਂ	কং	କଂ	కం	ಕಂ	–	കം	කං	کاں	–
kaṁ	कँ	કઁ	ਕੈਂ	কঁ	କଁ	–	–	–	–	–	(kān)	–

Chart depicting various Indian Languages written form (9)

It is found that there is one to one correspondence between what is written and what is spoken in Indian Language. Each character in Indian Language script has a correspondence to a sound of that language. In Indian Language , a consonant character is inherently bound with the vowel sound /a/ and is almost always pronounced with this vowel. This occurs at both word final and word middle positions. Since the letter to phone rules are straight forward in all Indian Languages, the syllabification rules are not trivial. Thus for Telugu language speech to text systems new rules are to be framed to break words into syllables, example C *VC* by grouping clusters based on heuristic analysis

III. PROCESS OF SPEECH TO TEXT CONVERSION

The field of Natural language processing is growing and has always been a good research area from past years. There are numerous applications of Natural Language Processing. Speech recognition is one of the most important applications of Natural Language Processing. Speech has always been the most important part of our day to day communication. We express our ideas through a specific language. Computers understand our language (natural language) by speech recognition. Speech or word by word recognition is the process of extracting the attributes of speech and classifying the same attributes with the prerecorded datasets. To recognize a word, word must be passed on to higher-level software for syntactic and semantic analysis. It is a technique of pattern matching, where audio signals are tested and framed into phonetics (number of words, phrases and sentences). To perform such task one needs to record a voice sample and then convert this voice sample into wav format. Spectrum based parameters are obtained when a word is recognized.

Speech or word recognition is the process of extracting the features of speech and classifying the same features with pre recorded data sets. The recognition of the word requires a higher level of software for its both syntactic and semantic analysis, by using the pattern matching where these audio signals are tested and its phonetics are framed.

The Microphone converts the Audio signal variations into electrical signals on which the Speech to text processing is done.
The main tasks involved in the Speech to Text conversion are
1. Speech Recognition
2. Speech Signal Enhancement
3. Pattern Recognition
4. Text Formation

IV. TECHNIQUES USED IN SPEECH TO TEXT CONVERSIONS

The above process are implemented by various techniques, and each have their individual strengths and challenges which are discussed as follows

a. Hidden Markov Model provides an elegant statistical framework for modeling speech patterns and is the most widely used technique. As a consequence , almost all Large Vocabulary continuous speech recognition (LVCSR) systems are based on HMMs The speech recognition system implemented using Hidden Markov Model (HMM) for representing speech sounds consists of a number of states, each of which is associated with a probability density function. The parameters of a HMM comprises of the parameters of the set of probability density functions, and a transition matrix that contains the probability of transition between states. The MFCC feature vectors extracted from speech signals and their associated transcriptions are used to estimate the parameters of HMMs. This process is called ASR system training.

The core of all speech recognition systems consists of a set of statistical models representing various sounds of the language to be recognized. (2) Since speech has temporal structure and can be encoded as a sequence of spectral vectors spanning the audio frequency range, the hidden Markov Model provides a natural framework for constructing such models.

The core of all speech recognition systems consists of a set of statistical models representing the various sounds of the language to be recognized. Since speech has temporal structure and can be encoded as a sequence of spectral vectors spanning the audio frequency range, the hidden Markov model (HMM) provides a natural framework for constructing such models.

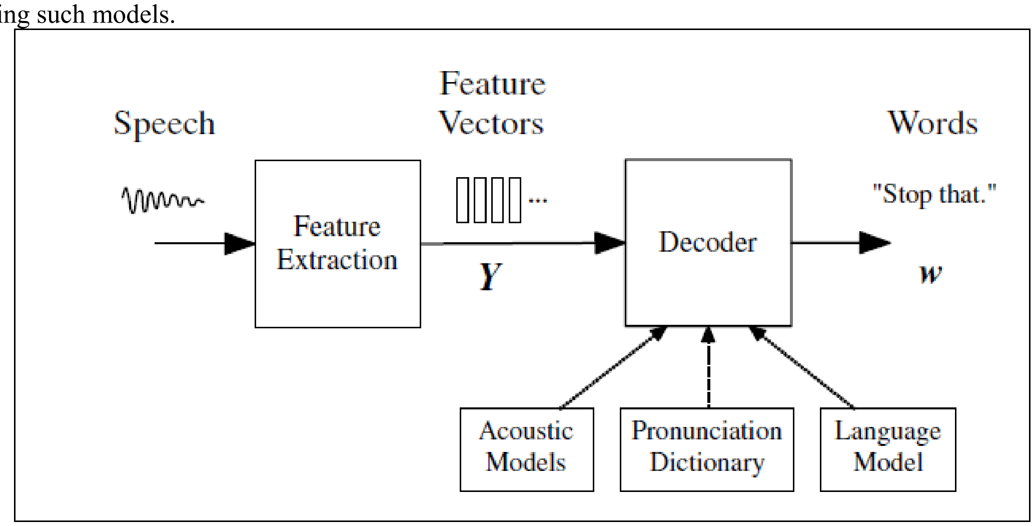

Fig. 1. Architecture of a HMM based speech recognizer (Adapted)

The Architecture of the Hidden Markov Models (HMM) based model.

The Hidden Markov Model (HMM) is a popular statistical tool for modeling a wide range of time series data. The important components of a large vocabulary continuous speech recognizer are shown in Fig.1. The input audio waveform from a microphone is converted into a sequence of fixed size acoustic vectors called feature extraction. The decoder then attempts to find the sequence of words, which is most likely to have generated by the decoder.

The likelihood is determined by an acoustic model and language model. The basic unit of sound represented by the acoustic model is the phone. About 40 such phones are required for English. Acoustic model is synthesized by concatenating phone models to make words as defined by a pronunciation dictionary. The parameters of these phone models are estimated from training data consisting of speech waveforms and their orthographic transcriptions. Some modern decoders generates compact representations.

One drawback of HMM is the various conditional independence assumptions imposed by the Markov Models. These assumptions essentially state that each speech frame is independent of its neighbors[(3)]

b. Artificial Neural Networks (ANN) is a computer system process used for speech to text conversion. It is inspired from the organization of cells of a human brain. Multilayer perceptrons (MLP) are the best studied class of ANN frequently applied in speech recognition. They have layered feed forward architecture with an input layer , zero or more hidden layers and as an output layer. In speech recognition using MLP, it is possible to present all at once , the acoustic vectors of a speech unit i.e. phoneme of the word at the input layer and to detect the most probable speech unit at the output layer by determining the output neurons with the highest activation.

V. LATEST TRENDS - LITERATURE SURVEY

1. M.S.E Langarani[10] et al (2012) in their paper reported that by incorporating the phase information in the enhancement process, the quality and intelligibility of speech signal are improved. In their investigations, the minimum mean-square error short-time spectral amplitude and MMSE log-spectral amplitude methods were used to estimate the magnitude spectrum of speech signal. By conducting six classes of experiments, it is shown that by taking the phase information into account, overall SNR and PESQ measures are improved. In other experiments, they explained the results of speech enhancement system are fed into a speech recognition system and it is shown that by incorporating accurate phase information, the intelligibility of the speech signal is improved. This is approach is different from that the majority of speech enhancement methods which perform noise removal in spectral domain and construct the enhanced speech signal from the estimated magnitude of clean speech and the phase of the noisy speech. Since speech enhancement methods aim to improve the quality and/or intelligibility of speech signal, but speech quality improvement does not necessarily result in improvement of intelligibility. Typically, single-channel noise reduction methods process the signal in Short Time Fourier Transform (STFT) domain. These methods focus on modifying the magnitude spectrum and do not alter the phase spectra. As these methods are based on a Gaussian model and the DFT coefficients of their noise reduction filters are real, they use the phase obtained from the DFT coefficients of the noisy speech signal as the phase of the enhanced signal. Nevertheless, it is known that considering phase information in estimation of clean signal improves the quality of the enhanced speech signal.

2. Jinkyu Lee11 et al (2012) has analyzed the impact of various preprocessing modules to improve the performance of automatic speech recognition system (ASR) in noisy environment. Though the performance of automatic speech recognition system has been increased significantly, the application area of the system is still limited to relatively clean environment because of performance degradation in adverse condition. The main reason can be found from the mismatch between training and testing data. Authors studied various methods to reduce the mismatch such as the ones designed in the signal domain, feature domain, and model domain. and they have selected the various state-of-the-art algorithms designed in the signal domain and feature domain, and evaluated thoroughly their performances in various noise conditions. Since the enhancement has been directly made to the features that are actually used for recognition, it was found that the feature domain approach is more appropriate than the signal domain approach. and proven through experimental results that the noise reduction in the feature domain gives the best performance.

3. Ryoji Miyahara12 and Sugiyama (2016) proposed gain relaxation in signal enhancement designed for speech recognition with an unaware local noise source. An attention is drawn to a new performance degradation problem in signal enhancement combined with automatic speech recognition (ASR), which is encountered in real products with an unaware noise source. This is eventually a new technique. Gain relaxation, as a solution, selectively applies softer enhancement of a target signal to eliminate potential degradation in speech recognition caused by small undesirable distortion in the target signal components. Evaluation of directional interference suppression with signals recorded by a commercial PC (personal computer) demonstrates that signal enhancement over the input is achieved without sacrificing the performance for clean speech.

4. Automatic speech recognition (ASR) is the process and the related technology for converting the speech signal into its corresponding sequence of words or other linguistic entities by means of algorithms implemented in a device, a computer, or computer clusters. **Jinyu Li[13], Li Deng, Yjfan Gong of Microsoft Corporation USA and Reinhold Haeb - Umbach of Univ of Paderborn, Germany** have worked on the consumer-centric applications, such as voice search and voice interaction with mobile devices and home entertainment systems, increasingly require automatic speech recognition (ASR) to be robust to the full range of real-world noise and other acoustic distorting conditions. Explained that Despite its practical importance, however, the inherent links between and distinctions among the myriad of methods for noise-robust ASR have yet to be carefully studied in order to advance the field further. To this end, it is critical to establish a solid, consistent, and common mathematical foundation for noise-robust ASR, which is lacking at present. In the article on **An Overview of Noise-Robust Automatic Speech Recognition[13] (2014),**they provided an effort to fill this gap and to provide a thorough overview of modern noise-robust techniques for ASR developed over the past 30 years. They have emphasized methods that are proven to be successful and that are likely to sustain or expand their future applicability. The authors considered all new parameters, keeping only those old parameters which are relevant. They have analyzed and categorized a wide range of noise-robust techniques using five different criteria: 1) feature-domain vs. model-domain processing, 2) the use of prior knowledge about the acoustic environment distortion, 3) the use of explicit environment-distortion models, 4) deterministic vs. uncertainty processing, and 5) the use of acoustic models trained jointly with the same feature enhancement or model adaptation process used in the testing stage. With this taxonomy-oriented review, we equip the reader with the insight to choose among techniques and with the awareness of the performance-complexity tradeoffs. The pros and cons of using different noise-robust ASR techniques in practical application scenarios are provided as a guide to interested practitioners. The current challenges and future research directions in this field is also carefully analyzed. Historically, ASR applications have included voice dialing, call routing, interactive voice response, data entry and dictation, voice command and control, structured document creation (e.g., medical and legal transcriptions), appliance control by voice, computer-aided language learning, content-based spoken audio search, and robotics. More recently, with the exponential growth of big data and computing power, ASR technology has advanced to the stage where more challenging applications are becoming a reality. Examples are voice search and interactions with mobile devices (e.g., Siri on iPhone, Bing voice search on winPhone, and Google Now on Android), voice control in home entertainment systems (e.g., Kinect on xBox), and various speech-centric information processing applications capitalizing on downstream processing of ASR outputs [3]. For such large-scale, real-world applications, noise robustness is becoming an increasingly important core technology since ASR needs to work in much more difficult acoustic environments than in the past.

5. The hybrid deep neural network (DNN)-hidden Markov model (HMM) has shown to significant improvement in speech recognition performance over the conventional Gaussian mixture model (GMM)-HMM. The performance improvement is partially attributed to the ability of the DNN to model complex correlations in speech features. **Ossama Abdel - Hamid[14] et al (2014)** in their paper shown that further error rate reduction can be obtained by using convolutional neural networks (CNNs). They experimented by presenting a concise description of the basic CNN and explained how it can be used for speech recognition. They further propose a limited-weight-sharing scheme that can better model speech features. The special structure such as local connectivity, weight sharing, and pooling in CNNs exhibits some degree of invariance to small shifts of speech features along the frequency axis, which is important to deal with speaker and environment variations. Experimental results shown that CNNs reduce the error rate by 6%-10% compared with DNNs on the TIMIT phone recognition and the voice search large vocabulary speech recognition tasks.

6. Aneeja Gunnam15 and B. Yagnanarayana have proposed a method for extracting the fundamental frequency (fo) from degraded speech signals using single frequency filtering (SFF) approach. The single frequency filtering of frequency-shifted speech signal gives high signal-to-noise ratio (SNR) segments at some frequencies and hence the SFF approach can be exploited for fo extraction using autocorrelation function of those segments. Since the fo is computed from the envelope of a single frequency component of the signal, the vocal tract resonances do not affect the fo extraction. The use of the high SNR frequency component in a given segment helps in overcoming the effects of degradations in the speech signal, without explicitly estimating the characteristics of noise. The proposed method of fo extraction is shown to give better performance for several types of real and simulated degradations, in comparison with some of the methods reported recently in the literature.

7. Abu Shafin Mohammed Mahdee Jameel16 et al (2016) proposed in their paper, a noise robust formant frequency estimation scheme based on a spectral model matching algorithm. They have Considered the vocal tract as an autoregressive system, a spectral model of repeated autocorrelation function (RACF) of band limited speech signal is proposed. It is shown that because of repeated autocorrelation operation on band-limited signal, the proposed model can exhibit prominent formant characteristics. First from given noisy speech observations, an adaptive band selection criterion is developed. Next, on each resulting band-limited noisy speech signal, a repeated autocorrelation operation is carried out, which not only reduces the effect of noise but also strengthens the dominant poles corresponding to the formant frequencies. Finally, spectrum of the RACF is computed and instead of direct spectral peak picking, a model fitting scheme is introduced to find out model parameters which lead to formant estimation. They tested the proposed algorithm on natural vowels as well as some naturally spoken sentences in the presence of different environmental noises. They found that the proposed scheme provides better formant estimation accuracy in comparison to some of the existing methods at low levels of signal-to-noise ratio.

8. Simon Receveur, Robin.W. Tim Fingscheidt17 -(2016) have reported that the Performance of automatic speech recognition (ASR) systems can significantly be improved by integrating further sources of information such as additional modalities, or acoustic channels, or acoustic models. Given the arising problem of information fusion, striking parallels to problems in digital communications are exhibited, where the discovery of the turbo codes by Berrou et al. was a groundbreaking innovation. They discussed in this paper the ways how to successfully apply the turbo principle to the domain of ASR and thereby provide solutions to the above mentioned information fusion problem. Experiment were done in four stages i.e : First, reviewing the turbo decoding forward-backward algorithm (FBA), giving detailed insights into turbo ASR, and providing a new interpretation and formulation of the so-called extrinsic information being passed between the recognizers. Second, presenting a real-time capable turbo-decoding Viterbi algorithm suitable for practical information fusion and recognition tasks. Then presenting simulation results for a multimodal example of information fusion. Finally, prove the suitability of both our turbo FBA and turbo Viterbi algorithm also for a single-channel multimodel recognition task obtained by using two acoustic feature extraction methods. On a small vocabulary task (challenging, since spelling is included), the designed proposed turbo ASR approach outperforms even the best reference system on average over all SNR conditions and investigated noise types by a relative word error rate (WER) reduction of 22.4% (audio-visual task) and 18.2% (audio-only task), respectively.

9. K.M. Shivakumar18 et al - (January 2017) investigated the complex problem of speech to text conversion of Kannada Language and proposed a novel Kannada Automated Speech to Text conversion System (ASTC). Natural Language Processing is an area under artificial Intelligence intended to accept process and manipulate the human language. It helps to model cognitive behavior of human mind to process natural language. The NLP is used to identify the natural language text and speech. In this process the mathematical, computational and linguistic knowledge has to be integrated to develop an application which serves as an aid to human beings in improving their task of perceiving their day to day information. NLP is used in Speech, Text, Sentiment Analysis and other applications. Application of Machine Learning Algorithm on archived natural language data such as mono-lingual or bi-lingual text or speech corpus acquires the techniques to process the natural language. Processing a natural language involves identifying a given language font, word, phrase or a sentence. This task opens a wide avenue of research to carryout in developing language processing tools such as dictionaries, wordnet, parts of speech taggers, morphological analyzers, machine translation systems and automatic speech recognizers. Speech processing is one of the application area of NLP which involves encoding and decoding the audio signals uttered by human being to produce a letter, word or sentence. Speech Processing uses Hidden Markov Models to extract features and other statistical models. These models are used to encode speech to its textual form and used in decoding to represent in textual form. Speech Recognition Systems which now rely on cache language models. These models extract the language pattern and the order of arrival of word or speech sequences and classify them based on their probability of occurrence in real time. These probabilistic estimations produce more accurate results when we are testing with unknown data or speech sequence. This was implemented for the Kannada Language by training and test the Speech Processing System using CMUSphinx framework. CMU Sphinx is dynamic in nature with support for other languages along with English. The Acoustic model for Kannada speech with 1000 general spoken sentences and tested 150 sentences were tested. The system was built utilizing features available in CMU Sphinx, thus showcasing the conceivable flexibility of this framework for Kannada voice to text conversion. In this paper, Kannada sentences with four to ten word length is researched. The speech conversion system permits ordinary people to speak to the computer in order to retrieve information in textual form. The number of alphabets in Kannada are 52. The system investigates extensibility of recognizing all letters and morphological variants of spoken Kannada words. Speech to Text Conversion had a wide range of application in human language processing. It is useful for business people to business meeting notes. Speaker identification and for deaf people is helpful to know the information.

10. Neha Sharma19 and Shipra Sardana (2016) created a real time speech recognition system that was tested in real time noiseous environment. We used the design of a bidirectional nonstationary Kalman filter to enhance the ability of this Real time speech recognition system. Bidirectional Kalman filter has been proved to be the best noise estimator in nonstationary noiseous environment. Real time speech to text conversion system introduces conversion of the uttered words instantly after the utterance. The purpose of this project was to introduce a new speech recognition system that is computationally simple and more robust to noise than the HMM based speech recognition system. We have used our own created database for its flexibility and TIDIGIT database for its accuracy comparison with the HMM based speech recognition system. MFCC features of speech sample were calculated and words were distinguished according to the feature matching of each sampled word. System was tested in different noise conditions and we obtained overall word accuracy of 90%.

VI. CONCLUSION

Speech Recognition is the procedure of extracting essential information from input speech signal to make accurate decision about the corresponding text[20]. Speech signal conveys very rich information, such as speaker information, linguistic information which has inspired many researchers to develop the system that automatically process the speech e.g. speech enhancement, speech synthesis, speech compression, speaker recognition, speech recognition and verification. Speech recognition can be further classified as speaker dependent and speaker independent. Human voice instructions are followed and understood by computers. with the help of speech recognition techniques. i.e. it acts as good interface for human computer interaction.

Various papers and abstracts were studied and found that the HMM and ANN models are found to be more used in the process of real time speech to text conversion. Comparatively to the HMM, Artificial Neural Network and Deep Neural Network is the latest technology being used. Gain Relaxation technique was also been referred as a increasing tool for accuracy of ASR. urther some authors found the spectral estimation as the performance increasers where the vocal tract considered as as an autoregressive system, a spectral model of repeated autocorrelation function (RACF) of band limited speech signal is proposed. Regarding the mismatch in Noisy environments feature domain approach is more appropriate than the signal domain approach were also reported.

Major speech recognition technologies are designed for English language. Maximum work has been done in the English speech to text conversions. Standard Tool Kits were used. standard word lists were also been developed for English language. It has found that European languages have reached a finer level of conversion, still much work is needed to be done for Indian Languages.

India is a vast country with different languages are spoken every 50 Kms and a need for the native language processing is important. and as digitization is growing with fast track, due to English language barrier, the illiterate rural communities or educationally under-privileged people, the disabled in speech and hearing peoples are being kept away of computer technology. If the processing of computer technology in native language is made possible i.e. if computer technologies can understand the native language then it will be easy to use computer technologies for illiterate people, people from rural communities or educationally under-privileged. Further Telugu is a prominent Language in India. There are two Telugu speaking states. Hence there is a need to develop the system for Telugu language.

Indian Languages are phonetic based and the both HMM and ANN can both give best accuracy in the speech conversion. The HMM and ANN models can be applied for the speech recognition for the Telugu language also. The speech pattern recognition pertaining to Telugu language consisting different 62 alphabets (7) including Consonants, Vowels and numbers. For conversion it is needed to understand its various fonts phonemes etc and translation needs to be done using the HMM or ANM models. There are various techniques available for the feature Extraction, Fundamental frequency extraction etc using such as single frequency filtering technique, which find out the F_o the fundamental frequency from the degraded speech and robust latest technologies like The hybrid deep neural network (DNN)-hidden Markov model (HMM) In this process the different patterns of the Telugu font like frequency, pitch etc are identified and studied. Apart from this some studies proposed gain relaxation in signal enhancement designed for speech recognition with an unaware local noise source. An attention is drawn to a new performance degradation problem in signal enhancement combined with automatic speech recognition (ASR), which is encountered in real products with an unaware noise source. This is eventually a new technique. thus by using these techniques and models it is able to develop a system in which when user speaks any alphabets into the microphone the different patterns of the alphabets will be identified and it will be compared with the corresponding pattern stored in the standard phoneme databases and corresponding highest matching alphabet of Telugu language will be return in form of text on the screen.

Thus apart from the rural population , Speech and Hearing Impaired persons can learn better when both the Audio and Visual reinforcements are given simultaneously. This approach of real time translation onto the Telugu language will improve the learning of the Speech and Hearing Impaired Children.

REFERENCES

[1] Thomas .F.Quatieri -Discrete time speech signal processing : principles and Practice - (Page 79)

[2] Mark Gales and Steve Young - Cambridge University Engineering Department, Trumpington Street, Cambridge, "The Application of Hidden Markov Models in Speech Recognition"

[3] Mondher Frikha* , Ahmed Ben Hamida Advanced Technologies for Medical and Signals (ATMS) Research Unit, National School of Engineering of Sfax, B.P.W, Sfax ,Tunisia.A Comparitive Survey of ANN and Hybrid HMM/ANN Architectures for Robust Speech Recognition

[4] L. R. Rabiner “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,”, Proceedings of the IEEE, Vol. 77, No. 2, 1989, pp. 257-286.

[5] Gaurav, Devanesamoni Shakina Deiv, Gopal Krishna Sharma, Mahua Bhattacharya at IIIT - Gwalior "Development of Application Specific Continuous Speech Recognition" - Journal of Signal and Information Processing, 2012, 3, 394-401

[6] Mishra A., Biswas A., Chandra M., and Sharan N., “Robust Hindi Connected Digit Recognition” -, International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 4, No.2, June 2011.

[7] S P Kishore, Rohit Kumar and Rajeev Sangal, “A Data Driven Synthesis Approach For Indian Languages using Syllables as BasicUnit” -, in Proceedings of Intl. Conf. on NLP (ICON) 2002, pp. 311-316, Mumbai, India, 200.

[8] M. Siva Kumar et al "Text to Speech System for Telugu Language " Int. Journal of Engineering Research and Applications www.ijera.com ISSN: 2248-9622, Vol. 4, Issue 3(Version 1), March 2014, pp.913-917

[9] Adapted from Google Indian language chart

[10] M. S. E. Langarani ,H. Veisi, H. Sameti - Dept. of Comput. Eng., Sharif Univ. of Technol The effect of phase information in speech enhancement and speech recognition, Tehran, Iran ieee published in IEEE *Xplore*: 24 September 2012

[11] Jinkyu Lee, Soonho Baek, Hong-Goo Kang of Department of Electrical and Electronic Engineering Yonsei University, Seoul, Korea Signal and feature domain enhancement approaches for robust speech recognition **published in** IEEE *Xplore*: **03 April 2012**

[12] Ryoji Miyahara and Sugiyama - Internet Terminal Division, NEC Engineering, Ltd., Information and Media Processing Laboratories, NEC Corporation, 1753 Shimonumabe, Nakahara-ku, Kawasaki 211-8666, Japan in the paper Gain relaxation: A useful technique for signal enhancement with an unaware local noise source targeted at speech recognition - Published in Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference - May 2016

[13] Jinyu Li, Li Deng, Yjfan Gongof Microsoft Corporation USA and Reinhold Haeb - Umbach of Univ of Paderborn .An Overview of Noise-Robust Automatic Speech Recognition , Germany Published in IEEE/ACM Transactions on Audio, Speech, and Language Processing (Volume: 22, Issue: 4, April 2014)

[14] Ossama Abdel - Hamid,Abdel - Hui Jiang- York University, Toronto Canada, Rahman Mohamed, Gerald Penn -University of Toronto, Canada, Li Deng, Dong Yu Microsoft Research, Redmond" Convolutional Neural Networks for Speech Recognition" - Published in: IEEE/ACM Transactions on Audio, Speech, and Language Processing (Volume: 22, Issue: 10, Oct. 2014)

[15] Aneeja Gunnam of IIIT - Hyderabad and B. Yagnanarayana "Extraction of fundamental frequency from degraded speech using temporal envelopes at high SNR frequencies" - Published in:IEEE/ACM , Transactions on Audio, Speech and Language Processing (Volume: PP, Issue: 99) 8th February 2017

[16] Abu Shafin Mohammed Mahdee Jameel, Shaikh Anowarul Fattah, Rajib Goswami, Wei - Ping Zhu, M Omair Ahmad Noise Robust Formant Frequency Estimation Method Based on Spectral Model of Repeated Autocorrelation of Speech -. Published in - Published in:IEEE/ACM ,Transactions on Audio, Speech and Language Processing (Volume: PP, Issue: 99) November 2016

[17] Simon Receveur, Robin.W. Tim Fingscheidt, Institute for Communications Technology, Technische Universität Braunschweig, Braunschweig, Germany, Turbo Automatic Speech Recognition- . Published in: IEEE/ACM Transactions on Audio, Speech, and Language Processing (Volume: 24, Issue: 5, May 2016

[18] K. M. Shivakumar, K G Aravind, T V Anoop, from Amrita University, Mysuru, Deepa Gupta, Amrita University, Bengaluru , Kannada speech to text conversion using CMU Sphinx - Published in: IEEE Explore Inventive Computation Technologies (ICICT), International Conference on 26 January 2017

[19] Neha Sharma and Shipra Sardana - Chandigarh University, A real time speech to text conversion system using bidirectional Kalman filter in Matlab, Published in: Advances in Computing Communications and Informatics (ICACCI) 2016, International Conference on 3rd November 2016

[20] Yogita H Ghadage and Sushama D Shelke - Pune."Speech to text conversion for multilingual languages published in : Communication and Signal Processing (ICCSP), 2016 International Conference on 2 November 2016