Automatic Speech Recognition Systems for Regional Languages in India

Ravindra Parshuram Bachate, Ashok Sharma

ABSTRACT--- ***Speech recognition systems has made remarkable progress in last ¬few decades such as Siri, Google assistant, Cortana. For improving the automation in services of all sectors including medical, agriculture, voice dialling, directory services, education, automobile etc., ASR systems must be built for regional languages as most of the Indian population in not familiar with English. Lots of work is done for English language but not for regional languages in India. Developing ASR and ASU systems will change the scenario of current service sector. There are many challenges in building ASR system, Noise reduction is a one of the challenging and still unsolved parameters which affects a lot on performance of any ASR system. Basically, three models required for building any ASR systems- Language model, acoustic model and pronunciation model. In this paper, discussed various parameters affecting on building ASR systems, development of ASR systems, Tools and Techniques used for building an ASR system and research on regional languages ASR system. Deep Neural network (DNN) provides a better way of recognising a speech and accuracy is high.***

Keywords— ASR, Acoustic Model, DNN, Language Model, Noise Reduction and Pronunciation model

I. INTRODUCTION

Since ten decades, research in Automatic Speech Recognition System has made significant development such as Microsoft Translator, Google voice search, Apple's Siri etc. This system is now used with great interest in both industry and academia. [1]. In voice recognition and communication system, speech recognition is actively considered, and it has tremendous progress due to digital signal processing. Any speech recognition system consists of three fundamental steps- pre-processing, feature extraction and classification. In the last decade, cost of component required for speech processing systems such as digital processors and communication devices reduced with improvement in performance and accuracy. Surrounding and noise makes speech recognition system complicated as it affects a lot on performance and accuracy of the Automatic Speech Recognition System. Speech Recognition accuracy is used for performance measurement of speech recognition system [2].

Human uses a speech as one of the effective medium for communication. Speech is a digital signal made up of various components like time, amplitude and frequency. Because of this, in different frequency bands, transitions occur at different time. In simple word, we can define Automatic Speech Recognition (ASR) as a system which extracts, recognize and translate properties of speech using computer device. The purpose of the developing ASR system is to enable human real time interaction with a machine in natural language regardless of speech accent, speech noise, environment and size of vocabulary. ASR systems can be developed for recognizing isolated words, connected word and continuous speech. ASR systems has various industrial and academia applications such as telecommunication, for impaired hearing people, learning language for children's etc. [3].

Speech is a multi-component signal with varying time, frequency and amplitude. Due to this variability, transitions may occur at different times in different frequency bands. The technology developed for extraction, recognition, and translation of the speech characteristics by using the smart computerized device is called an Automatic Speech Recognition (ASR). The main purpose of ASR is to develop a technology which helps human to interact with the machine in our natural language in real time environment regardless of vocabulary size, noise, speech characteristics or accent[51, 52]. ASR can be implemented for isolated word recognition, connected word recognition, and continuous speech recognition.

This paper is divided into five sections. First section discusses introduction to Speech Recognition System, second section discusses overview of Automatic Speech Recognition System, in third section, literature review, fourth section discusses about ASR systems for Indian languages and paper is concluded in fifth section.

II. OVERVIEW OF ASR SYSTEM

This section covers points related with ASR system such as history of speech recognition, types of ASR system, architecture of ASR system, challenges in ASR System, feature extraction techniques and pattern recognition or classifier. Figure 1 describes ASR types, type of speech and features extraction and pattern recognition techniques.

Revised Manuscript Received on July 10, 2019.

Ravindra Parshuram Bachate, Research Scholar, School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India. (E-mail: bachateravi@gmail.com)

Dr. Ashok Sharma, Associate Professor, School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India. (E-mail: drashoksharma@hotmail.com)

Retrieval Number: B11080782S319/19©BEIESP
DOI : 10.35940/ijrte.B1108.0782S319

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

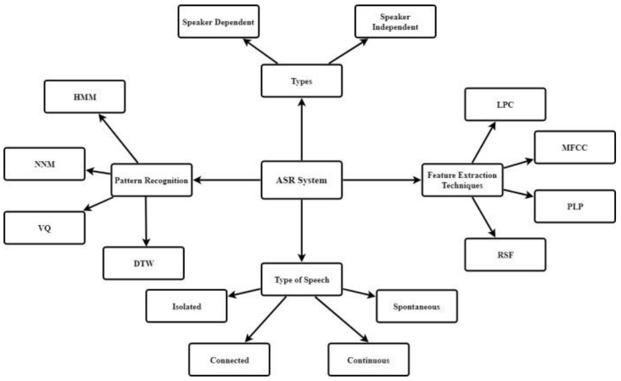

Figure 1 Overview of ASR system

A. History of Speech Recognition

Speaking is a primary media of communication for the people. Researchers are always trying to make machineintelligent and human interactive. Radio REX was first ASR system was introduced in 1922 which recognizes few words only. In 1930, at Bell laboratories, Homer Dudley proposed a system for speech analysis and synthesis which was the beginning of speech recognition system [4]. The continuous efforts to build an Automatic Speech Recognition (ASR) are going on since last 10 decades. The development in speech recognition system is addressed in figure 2.

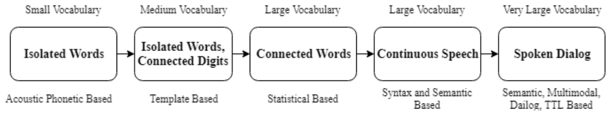

Figure 2 ASR system Development over ten decades [4]

The ASR system development starts with identifying an isolated word with very small vocabulary. This model was designed only to recognize few isolated words. This model was based on acoustic and phonetic having very small vocabulary. Later Phonetic Typewriter was introduced in 1957 by H. Belar et al. [5]. Later the speech recognition system developed were based on the template. These systems were able to recognize isolated words and connected digits. These ASR systems uses medium size vocabulary. After this, a statistical based ASR systems were introduced and these ASR systems having large vocabulary size. Due to this, these systems can recognize connected words. Later, the ASR system which was developed are based on the statistical models. These systems could recognize the connected words and having large vocabulary size. But just recognizing connected words was not a satisfactory solution for ASR system. That is the reason, a new syntax and semantic based ASR system was developed having large vocabulary. After this an ASR system which supports spoken dialog were introduced. This system is based on semantic, multimodal dialog and TTS. This ASR system uses very large vocabulary for implementing the system.

B. Architecture of ASR System

Automatic Speech Recognition systems are built up with various models and phases described in Figure 3. Usually, all ASR systems are having three phases, signal processing, feature extraction and decoding. Again, decoding carried out its task with the help of acoustic model, language model and pronunciation model.

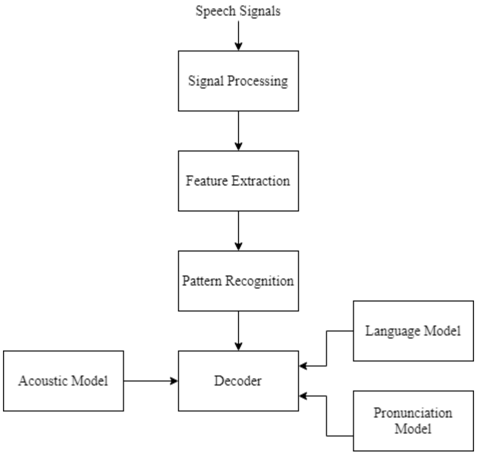

Figure 3 ASR System Architecture

Speech Signal is input to all ASR systems designed but the medium and type of speech may differ. Speech is a sequence of sounds with different properties associated with it. These sounds are converted into small segments called frame. Feature extraction processes these frames to extract features in terms of speech vectors. From these, useful characteristics are extracted which are useful for identifying a basic sound ‘a’, ’u’ and ‘m’. Speech recognition is like the pattern recognition system. Decoder basically uses three models mentioned in Figure 3 acoustic model, language model and pronunciation model for decoding the sounds and convert it into transcription.

C. Types of Speech

Basically, speech is classified into four types. Based on the type of speech, ASR system need to define. The type of speech is given below as-

1) Isolated Words

Isolated words speech having a significant utterance wait time between two words. Isolated word doesn’t mean only single word is accepted but it accepts only one word at a time. ASR system based on isolated word can be used in home atomization.

2) Connected Words

Connected words and isolated words are not having much difference. The only difference between these two types of speeches are isolated words having significant pause between two successive words whereas connected words having less wait time between two words.

3) Continuous Speech

Human’s natural speech considered as a continuous speech in developing speech recognition system. It is a challenging task to recognize continuous speech due to various aspects of speech. Difficult task in continuous speech is to identify the utterance boundaries between the two successive words.

Retrieval Number: B11080782S319/19©BEIESP
DOI : 10.35940/ijrte.B1108.0782S319

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

4) Spontaneous Speech

Spontaneous speech is like continuous speech, but it is having ability to recognize the words even though being uttered together. Spontaneous speech recognition system has an ability to tackle various kind of natural speeches. It is a challenging task to develop automatic speech recognition system for spontaneous speech due to variation in speech pronunciation, speech noise, pitch etc.

D. Challenges in Speech Recognition

Even though speech recognition work is going on since 1920, still there are lot of challenges in speech recognition system which remains unsolved [6]. Issues which effects on accuracy of speech are discussed below-

1) Style of speech

Style of speech depends on various attributes like Voice tone, accents, rate of speech, voice pitch and phoneme production [6]. Voice tone may be shouted, normal and quite. Accent for same language varies person to person and region to region. Phoneme production is depending on what type of speech it is. Speech may be of type isolated words, connected words, continuous speech or spontaneous speech. Sometimes tone pitch also causes for speech variability and it is difficult for ASR systems to recognize the words.

2) Environment

Environment is one of the most difficult hurdles in speech recognition system. Environment can be expressed through background noise, room acoustics and channel conditions. These parameters add noise to signal and make voice noisy.

3) Speaker characteristics

Variability in speech depends on the characteristics of the speaker. Speaker characteristics includes speaker’s age, sex and variation in articulation [6]. Variation in articulation includes the mental state of speaker, stress, emotions etc.

4) Language Characteristics

Developing ASR system for different languages needs different approaches as each language has its own structure. Every language has their own grammar suite and phonetical utterances.

E. Feature Extraction Techniques

Feature extraction is one of the significant steps used to develop an ASR system as it is used to extract the features from speech signal. There are four techniques listed and described below used for the feature extraction.

1) Linear Predictive Coding (LPC)

In linear predictive coding, speech frames are examined to produce speech vector. It is mandatory for speech signal to go through pre-emphasizer to extract feature from input speech signal. Pre-emphasizer works when speech signals are converted into speech frames or blocks. Next step after pre-emphasizer is windowing. In windowing, signal disruption is reduced at both the end of the frame i.e. at start and end of the frame. Next to windowing, windowed frames are correlated which results in order of LPC analysis. Based on the LPC analysis, LPC coefficients are generated.

2) Perceptual Linear Prediction (PLP)

Perceptual linear prediction model was introduced by Hermansky. It uses perception of hearing psychophysics for examining the speech signal. PLP changes SR rate as it does not accept speech signals inappropriate information. It is like linear prediction coding but modifies SR rate. It is a better approach of feature extraction compare to LPC as it focuses on human speech signals.

3) Mel Frequency Cepstral Coefficient (MFCC)

MFCC is most widely used feature extraction technique for speaker as well as speech signals. MFCC estimates human speech signals more accurately compare to other feature extraction systems or algorithms. MFCC computes features from signal frames based on short term analysis. Signal disruption is reduced in speech signal frames by applying hamming windows. Then Mel filter bank is produced by applying Discrete Fourier Transform (DFT).

$$m = 2595 \log 10 \; (1 + f700)$$

4) Relative Spectral Filtering (RSF)

Relative spectral filtering removes unnecessary noise from speech signals. RASTA technique is used for reducing noise effect and surrounding noise. Feature coefficient passed through RASTA filter.

F. Classifier / Pattern Recognition Approach

Classifiers algorithms are used for building different models in speech recognition system. There are four main approaches used in designing models- Hidden Markov Model (HMM), Neural Network Model (NNM), Vector Quantization (VQ) and Dynamic Time Warping (DTW).

1) HMM

Hidden Markov model is one of the popular models which follows stochastic approach to cope with incomplete information in speech. HMM is characterized by a finite state Markov model and a set of output distributions [7]. In HMM, temporal variability is a transition variability whereas spectral variability is a distribution model.

2) NNM

Neural Network Models are having capability to solve more complicated recognition tasks but as vocabulary size increases performance goes down compare to HMM [8]. Neural network allows ASR system to recognize complex speech such as spontaneous or connected as it is difficult to identify boundaries between two successive connected words.

3) VQ

In vector quantization, functions of probability density are modelled using prototype vectors distribution. Vector Quantization (VQ) makes clusters of set of vectors having similar characteristics. It works like K-means algorithm where one point represents center point.

4) DTW

Dynamic Time Warping (DTW) approach finds the optimal warping between 2 time series. In DTW approach, Dynamic Programming (DP) is used which helps us to find out the lowest distance path to reduce the computation amount. Time synchronization is used for dynamic programming where each column of time matrix is taken into consideration.

Types of ASR Systems

Based on the speaker mode, ASR system is divided into two types – speaker dependent ASR system and speaker independent ASR system. Approaches for designing these ASR systems are different and its applications are different. Basically, speaker dependent ASR systems are developed for personal use whereas speaker independent systems are developed for general use. These two types of ASR systems are described and compared in below given table I.

Table I. Categories of ASR Systems

Parameters	Categories	
	Speaker Dependent	Speaker Independent
Trained by	Individual	Anyone
Accuracy	High Accuracy	Less compared to speaker dependent
Response	Accurate only for individual who trained the system	Good for anyone as it is trained by many speakers.
Scope	Used for developing personal software's	Used for industrial applications like tele services

III. RELATED WORK

Y. Murase et al. [9] discussed various techniques, Dialog state tracking, Knowledge base, Knowledge graph and associative knowledge inference. Based on the knowledge graph, feature vector is drawn with respect to associative knowledge. In this paper, proposed systems effectiveness which uses neural network is shown through experiments. Based on the results received, it is proposed that for DST, feature vector is a best solution.

Algorithm used in this paper are RNN (required neural network), CNN based approaches, AKFVS (Association knowledge feature vectors) and FCNN (Fully connected neural network). The inference method proposed for creating associative feature, it can be used for understanding spoken language aspects.

J. Guo et al. [10] proposed a speaker verification system for short utterances based on deep neural network. In this paper, male and female speaker's utterances are evaluated using DNN1 and DNN2, proposed system shows significant performance as compare to other techniques used in the paper. The performance of i-vector speaker based on GMM-DNN increased as utterance evaluation duration increases.

Ming Li et al. [11] proposed end-to-end deep learning methods achieve superior performance at the segment level, but not at the person-level. Deep learning methods for getting the best improvement in performance at segment level. Accuracy at person and segment level is improved based on the cross-validation technique. For finding patterns in speech, proposed system gives a superior result not at person level but at segment level.

Sibo Tong et al. [12] introduced cross-lingual acoustic model based on Connectionist Temporal Classification (CTC). This proposed model works efficiently and effectively when the phonemes are overlapped. It is also work in same way for new phonemes as it works for overlapped phonemes. Results in this paper shows proposed CTC based approach gives a significant performance compare to DNN/HMM approach with limited data set. In this paper, total three CTC based models- acoustic, multilingual and learning hidden unit contribution model

J. Takamya et al. [13] developed a dialogue break down detector for cluster annotators. It groups independent detectors trained by each cluster. The results shown in papers proposed system based on CRF baseline outperforms for DBDC3 Japanese tasks. C Bechikh Ali et al. [14] presented an empirical evaluation of compounds indexing for Turkish texts. Due to the pattern's ambiguity, Patterns 9, 10, 11 and 12 excluded from experiments. Indeed, a deep linguistics processing is needed to extract these compound types. Turkish Milliyet IR dataset using compounds to index documents and queries showed significant improvement.

Sidorov et al. [15] performed experiments which defines hypothesis stating that we can represent sentences by linking amount of information within sentences in terms of words and the context learned by word embedding's. In this paper, experiments are performed on STS tasks i.e. word embedding average and BoW. Experiments gives best results for sentence representation. After performing operations on STS tasks, SICK data set is evaluated which gives outstanding performance with a correlation of r¼ 0:724. It is good as it is a difficult task to measure similarity for semantic on dataset using unsupervised sentence representation.

Z. Wang et al. [16] implemented different mechanisms to develop custom service dialogs by taking help of information related to dialog history and response generation procedure external knowledge. Application which are developed based on response generation external knowledge and dialog history gives valuable and responsive results compare to baseline method. In this proposed system, still more attention should be given on deficiency of background information and timeliness inconsistency.

Tomas Kincl et al. [17] proposed an independent sentiment analysis model for language and domain which is based on character n-grams. It is tested, and it outperforms when it is trained on paired combination of domains. Y. Shi et al. [18] reviewed interaction between background noise and speech variations. This paper results shows that compare to CI listener, NH listener gives better performance for steady state noise. Both the listeners' performance depends on background noise and speech rate. But it is observed that speaking rate affects significantly on CI listener than NH listener. In this paper, two algorithms are used, Speech Reception Threshold (SRT) and long-term Root Mean Square (RMS).

Yen An-zi et al. [19] introduced an English Chinese bilingual word representation. In this paper, author investigated the various issues associated with word alignment. It also focuses on bilingual representation of words. Paper explored various non-alignment and alignment approaches to generate contexts for training the Skip-gram model. Algorithms used in this paper are Cross-lingual word semantic relatedness, Cross-lingual analogy reasoning and bilingual dictionary induction. Fei Chen et al. [20] contributed in cochlear implant speech perception. This paper describes merits about bilateral CI hearing. Proposed system helps in improving the perception about sentences which are interrupted segmentally.

Sangkeun Jung [21] proposed a system to learn distributed vectors of a semantic frame based on the neural network. This proposed system used two characteristics reconstruction and embedded correspondence for representation of meaningful and valid distributed semantic. Proposed system minimizes the distance between semantic frame reader and semantic vector. Semantic vectors which are equivalent placed close in the vector space.

Emre Yılmaz et al. [22] developed a semi supervised acoustic model for speech recognition. This developed model assigns language label to speech signals. This system also gives labels to speaker which results in best automatic annotation. This helps to design acoustic model having better accuracy and performance. The acoustic model designed here is designed for bilingual. Yanmin Quian et al. [23] proposed a single channel multi talker ASR system. In this paper, streams of multiple speech are separated by using permutation invariant. Separate feature streams can be obtained by implementing separation module on front end feature.

Ewald van der Westhuizen et al. [24] proposed an ASR system for four South African language pairs. For ASR system implementation, code switching technique is used which is a spontaneous and natural. Due to the globalization, code switching becoming popular technique. Code switching provides excessive soothing for language model.

Moataz et al. [25] reviewed speech emotion recognition system. Last accuracy was 50% and it is extended to 90%. Algorithms discussed in this paper are Support vector machine, neural network and Gaussian mixture models. Sadavoki et al. [26] implemented Speech-to-Text and Speech-to-Speech Summarization for spontaneous speech. Speech summarization technology can be applied to any kind of speech document and is expected to play an important role in building various speech.

Prerana et al. [27] developed a voice recognition system which translates the speech to text. In this paper, two models are used- speaker dependent and speaker independent model. This system will allow the computer to translate voice request and dictation into text using MFCC and VQ techniques. Feature extraction and feature matching will be done using Mel Frequency Cepstral Coefficients and Vector Quantization technique. Simon et al. [28] demonstrated various accommodation levels which extracts dominance hierarchy for conversations. This paper focuses on speech understanding and affects of behavior of human on it. Human's emotion needs to consider for spoken language understanding.

Vincent et al. [29] analyzed various factors in robust speech recognition system. In this paper, three parameters-environment, data simulation and microphone are discussed. This paper uses MVDR and DNN algorithms for building automatic speech recognition technique. Neha S. et al. [30] implemented a bi-directional speech to text conversion system. This paper introduces a technique for speech recognition which uses Bidirectional nonstationary KALMAN FILTER algorithm for building ASR. This system is simple and robust than HMM based ASR system.

Ismo K. et al. [31] used spectral resolution enhancement for improving noise reduction in audio signals. This paper is focuses on two things to reduce the noise in audio signal-Spectral resolution enhancement and spectral domain processing. Algorithms used in this paper are Linear prediction (LR)-used for modelling audio signals, Finite Impulsive Response (FIR) filter, autoregressive (AR) model, Standard Spectral Subtraction (SSS) method and psychoacoustic spectral subtraction method -used for calculating ten masking thresholds. Resolution enhancement procedure is based on extrapolating the signal in the time domain. More accurate noise attenuation with less signal distortion can be achieved with increased spectral resolution. Mark et al. [32] developed a system for creating small vocabulary and automatic speech recognition. This ASR system requires small size training corpus. This proposed system gives 67% accuracy.

Thimmaraja et al. [33] developed and compared ASR system for Kannada language using Kaldi tool. In this paper, two main technologies are used Spectral Subtraction with Voice Activity Detection (SS-VAD) and Minimum Mean Square Error-Spectrum Power Estimator (MMSE-SPZC) based on Zero Crossing respectively. It is a difficult task to develop spoken query system to access any data which is implemented in this paper.

Mittal et al. [34] proposed and implemented a ASR system for Punjabi language under different acoustic conditions. The major limitation of context dependent untied model is the requirement of higher memory space. This system is specifically developed for mobile phones. Songfang H. et al. [35] introduced language models based on Hierarchical Bayesian for conversational speech. Parallel training models used in this proposed system which allow proposed system deal with large size corpus. Here, two type of language models are used - SMOOTHING N-GRAM and Pitman–Yor processes.

Deepali M. et al. [36] implemented an ASR system for Marathi numerals and techniques used in this system are Mel Frequency Cepstral Coefficient (MFCC) and Low pass Filter Zero Interpolation (LFZI). Here, 1000 words corpus is used and implemented using MATLAB.

IV. ASR SYSTEMS FOR INDIAN LANGUAGES & RESULTS

Since last decade, Indian researcher are started working on regional languages in India. Most of the Indian population is not able to understand English language as a result even though ASR based technologies developed that are not useful for most of the people of India. Total 22 languages are officially spoken in India. As research work on speech recognition systems for Indian languages going on, but still it is in developing stage and not used for the commercial applications like home automation etc. Below mentioned table describes the research work on Indian languages with methodology used.

Table II. ASR Systems for Indian Languages

Author	Language	Feature Extraction Methodology	Model/ Classifier	Accuracy
[33]	Kannada	*SS-VAD and MMSE-SPZC*	GMM	17.01% and 13.18% for noisy and enhanced speech respectively
[34]	Punjabi	MFCC	CI,CD-Untied CD-Tied, D_DelInterp	CD_Untied gives highest accuracy with rate 81%
[2]	Marathi	MFCC	DWT, LPC& ANN	78%
[36]	Marathi	MFCC & LFZI	SVM	-
[37]	Marathi	MFCC	HMM & GMM	80-90%
[38]	Punjabi	MFCC and GFCC	HMM-GMM & DNN-GMM	4-5% improved performance in DNN-GMM compare to HMM-GMM
[39]	Punjabi	MFCC & PLP	*N*-gram model, MPE	MFCC gives good results compare to PLP
[40]	Tamil	SS–NE, CSD-NE	FCM with EM-GMM	Accuracy improved from 1.2 to 4.4%
[41]	Assamese	MFCC	HMM, VQ and I-vector	90-100%
[42]	Chhatisgarhi	-	HMM,ANN and SVM	99.84 and 94.24% for isolated words recognition systems using ANN and SVM respectively
[43]	Hindi	MFCC	*Genonic HMM, Segmental HMM, Hybrid HMM*	-
[44]	Telugu	MFCC-GMM, Prosodic-NNC	*GMM*	88-77% accuracy
[45]	Hindi	MFCC	VQ-GMM	93% accuracy
[46]	Gujrati	MFCC	HMM	Accuracy 95.9% in lab and 95.1% in noisy environment.
[47]	Kannada	MFCC	MLLT	Maximum 90.05% accuracy achieved for Online Mandi Phoneme based Web application
[48]	Urdu	MFCC	Bidirectional LSTM, RNN	WER is 0.68
[49]	Bengali	MFCC	Bidirectional LSTM	98.9% classification accuracy
[50]	Dravidian	MFCC + SDC	ANN	Dravidian language classification accuracy - 73.6%,72%, 65.1% and 68.8% for Kannada, Malayalam Tamil and Telugu respectively

V. DISCUSSION AND CONCLUSION

A lot of work is done on ASR systems for English language but not for Indian languages. Since last decade, Indian researchers are focusing on developing ASR systems for Indian languages. But till date, no successful ASR systems are commercially available for most of the Indian languages. In this paper, we discussed brief history of ASR systems, architecture of ASR system, different feature extraction techniques, various approaches of pattern recognition, type of ASR systems and some research done till date on regional languages in India.

ACKNOWLEDGEMENT

I express true sense of gratitude towards my research supervisor Dr. Ashok Sharma, Associate Professor, School of Computer Science and Engineering, Lovely Professional University, Jalandhar, Punjab for his valuable suggestion during preparation of this research article. I am also very much thankful to my family and friends for providing me all kind of support required.

REFERENCES

1. S. Toshniwal et al., “Multilingual Speech Recognition With A Single End-To-End Model Toyota Technological Institute at Chicago,” pp. 4904–4908, 2018.
2. A. Narkhede, “Efficient Method for Isolated Marathi Digits Recognition using DWT and Soft Computing Techniques,” 2018 3rd Int. Conf. Internet Things Smart Innov. Usages, pp. 1–5, 2018.
3. N. D. Londhe, “Recognition for Chhattisgarhi,” 2018 5th Int. Conf. Signal Process. Integr. Networks, pp. 667–671, 2018.
4. B. H. Juang and L. R. Rabiner, “Automatic Speech Recognition – A Brief History of the Technology Development,” pp. 1–24.
5. H. O. ; H Belar, “Phonetic Typewriter,” IRE Trans. Audio, vol. 5, no. 4, pp. 90–95, 1957.
6. P. Sahu, M. Dua, and A. Kumar, “Challenges and Issues in Adopting Speech Recognition,” Springer, vol. 664, pp. 209–215, 2017.

7. S. J. Arora, “Automatic Speech Recognition : A Review Automatic Speech Recognition : A Review,” no. September, pp. 33–44, 2017.
8. S. K. Saksamudre and R. R. Deshmukh, “A Review on Different Approaches for Speech Recognition System,” no. September, 2015.
9. Y. Murase, Y. Koichiro, and S. Nakamura, “Associative knowledge feature vector inferred on external knowledge base for dialog state tracking I,” Comput. Speech Lang., vol. 54, pp. 1–16, 2019.
10. J. Guo et al., “Deep neural network based i-vector mapping for speaker verification using short utterances,” Speech Commun., 2018.
11. M. Li, D. Tang, J. Zeng, T. Zhou, and H. Zhu, “An Automated Assessment Framework for A typical Prosody and Stereotyped Idiosyncratic Phrases related to Autism Spectrum Disorder,” Comput. Speech Lang., 2018.
12. S. Tong and P. N. Garner, “Cross-lingual Adaptation of a CTC-based multilingual Acoustic Model,” 2018.
13. J. Takayama, E. Nomoto, and Y. Arase, “Dialogue breakdown D13X X D14X X detection D15X X robust to variations D16X X in D17X X annotators and dialogue D18X X D19X X systems,” vol. 54, 2019.
14. X. X. D. Bechikh and D. X. X. S. D. X. X, “Empirical evaluation D5X X of compounds indexing for Turkish texts,” Comput. Speech Lang., 2019.
15. S. D. X. X, “Unsupervised sentence representations as word information series : Revisiting TF À IDF,” 2019.
16. Z. Wang, Z. Wang, Y. Long, J. Wang, Z. Xu, and B. Wang, “Enhancing Generative Conversational Service Agents with Dialog History and External Knowledge,” Comput. Speech Lang., 2018.
17. M. Nov, “Improving sentiment analysis performance on morphologically rich languages : Language and domain independent approach I a,” vol. 56, pp. 36–51, 2019.
18. Y. Shi et al., “Interaction between speech variations and background noise on speech intelligibility by Mandarin-speaking cochlear implant patients,” Speech Commun., vol. 104, no. August, pp. 89–94, 2018.
19. Y. An-zi, H. Hen-hsen, and C. Hsin-hsi, “Learning English À Chinese D21X X bilingual word representations from sentence-aligned parallel corpus I,” 2019.
20. F. Chen and Y. Hu, “Segmental contributions to cochlear implant speech perception,” vol. 106, no. December 2018, pp. 79–84, 2019.
21. S. Jung, “Semantic Vector Learning for Natural Language Understanding,” Comput. Speech Lang., 2018.
22. E. Yılmaz, M. Mclaren, H. Van Den Heuvel, D. A. Van Leeuwen, E. Yılmaz, and M. Mclaren, “Semi-supervised acoustic model training for speech,” Speech Commun., 2018.
23. Y. Qian, X. Chang, and D. Yu, “Single-channel multi-talker speech recognition with permutation invariant training,” Speech Commun., vol. 104, no. July, pp. 1–11, 2018.
24. V. D. W. Dx and R. N. Dx, “Synthesised bigrams using word embeddings for code-switched ASR of four South African language pairs,” Comput. Speech Lang., 2018.
25. M. El, M. S. Kamel, and F. Karray, “Survey on speech emotion recognition : Features , classification schemes , and databases,” Pattern Recognit., vol. 44, no. 3, pp. 572–587, 2011.
26. S. Furui, T. Kikuchi, Y. Shinnaka, and C. Hori, “Speech-to-Text and Speech-to-Speech Summarization,” vol. 12, no. 4, pp. 401–408, 2004.
27. P. Das, K. Acharjee, P. Das, and V. Prasad, “VOICE RECOGNITION SYSTEM : SPEECH-TO-TEXT,” no. July, 2016.
28. S. F. Worgan and R. K. Moore, “Towards the detection of social dominance in dialogue,” Speech Commun., vol. 53, no. 9–10, pp. 1104–1114, 2011.
29. M. Electric, “An analysis of environment , microphone and data simulation mismatches in robust speech recognition,” 2016.
30. N. Sharma, “A REAL TIME SPEECH TO TEXT CONVERSION SYSTEM USING BIDIRECTIONAL KALMAN,” pp. 2353–2357, 2016.
31. I. Kauppinen and K. Roth, “Improved Noise Reduction in Audio Signals Using Spectral Resolution Enhancement With Time-Domain Signal Extrapolation,” vol. 13, no. 6, pp. 1210–1216, 2005.
32. M. S. Hawley et al., “A Voice-Input Voice-Output Communication Aid for People With Severe Speech Impairment,” vol. 21, no. 1, pp. 23–31, 2013.
33. T. Y. G, “Development and Comparison of ASR Models using Kaldi for Noisy and Enhanced Kannada Speech Data,” pp. 1832–1838, 2017.
34. P. Mittal and N. Singh, “Development and analysis of Punjabi ASR system for mobile phones under different acoustic models,” Int. J. Speech Technol., vol. 0, no. 0, p. 0, 2019.
35. S. Huang and S. Renals, “Hierarchical Bayesian Language Models for Conversational Speech Recognition,” vol. 18, no. 8, pp. 1941–1954, 2010.
36. D. Malewadi, “Development of Speech recognition technique for Marathi numerals using MFCC & LFZI algorithm.”
37. S. S. Chavan and S. M. Handore, “Speech Recognition using HTK Toolkit for Marathi Language,” 2017 IEEE Int. Conf. Power, Control. Signals Instrum. Eng., pp. 1591–1597, 2017.
38. V. Kadyan, A. Mantri, and R. K. A. Amitoj, “A comparative study of deep neural network based Punjabi-ASR system,” Int. J. Speech Technol., vol. 0, no. 0, p. 0, 2018.
39. J. G. A. N. Mishra, “Continuous Punjabi speech recognition model based on Kaldi ASR toolkit,” Int. J. Speech Technol., vol. 0, no. 0, p. 0, 2018.
40. M. K. M. K. R. S. Valarmathi, “Continuous Tamil Speech Recognition technique under non stationary noisy environments,” Int. J. Speech Technol., vol. 0, no. 0, p. 0, 2018.
41. S. Sruba, B. Sanjib, and K. Kalita, “Speech recognition with reference to Assamese language using novel fusion technique,” Int. J. Speech Technol., vol. 0, no. 0, p. 0, 2018.
42. N. D. Londhe and G. B. Kshirsagar, “Chhattisgarhi speech corpus for research and development in automatic speech recognition,” Int. J. Speech Technol., vol. 21, no. 2, pp. 193–210, 2018.
43. R. K. Aggarwal and M. Dave, “Integration of multiple acoustic and language models for improved Hindi speech recognition system,” Int. J. Speech Technol., vol. 15, no. 2, pp. 165–180, 2012.
44. K. Mannepalli, P. N. Sastry, and M. Suman, “MFCC-GMM based accent recognition system for Telugu speech signals,” Int. J. Speech Technol., vol. 19, no. 1, pp. 87–93, 2016.
45. U. G. Patil, S. D. Shirbahadurkar, and A. N. Paithane, “Automatic Speech Recognition of isolated words in Hindi language using MFCC,” Int. Conf. Comput. Anal. Secur. Trends, CAST 2016, pp. 433–438, 2017.
46. J. H. and D. B., “Speech Recognition System Architecture for Gujarati Language,” Int. J. Comput. Appl., vol. 138, no. 12, pp. 28–31, 2016.

47. T. G. Y. H. S. Jayanna, "A spoken query system for the agricultural commodity prices and weather information access in Kannada language," Int. J. Speech Technol., vol. 20, no. 3, pp. 635–644, 2017.
48. T. Zia and U. Zahid, "Long short-term memory recurrent neural network architectures for Urdu acoustic modeling," Int. J. Speech Technol., vol. 22, no. 1, pp. 21–30, 2019.
49. T. Bhowmik, S. Kumar, and D. Mandal, "Manner of articulation based Bengali phoneme classification," Int. J. Speech Technol., vol. 21, no. 2, pp. 233–250, 2018.
50. S. G. Koolagudi, A. Bharadwaj, and Y. V. S. Murthy, "Dravidian language classification from speech signal using spectral and prosodic features," Int. J. Speech Technol., vol. 20, no. 4, pp. 1005–1016, 2017.
51. Saranya, E., Sam, B.B., Sethuraman, R. "Speech to text user assistive agent system for impaired person" 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, ICSTM 2017
52. R.Subhashini and V. Jawahar Senthil Kumar, "A Framework for Efficient Information Retrieval using NLP Techniques ", Proceedings of the International Conferences on Advances in Communication Network and Computing, CNC 2011, CCIS 142, pp. 391–393, 2011, Springer-Verlag Berlin Heidelberg 2011,ACEEE, Bangalore.