Speech Recognition Methods Applied for Different Punjabi Language Accents: A Review

Ravinder Singh[1], Dr.Anand Sharma[2]

M.Phil (CA) Scholar[1], Assistant Professor[2]

UCCA, Guru Kashi University, TalwandiSabo, Bhatinda (India)

ABSTRACT

Punjabi language is widely used in Punjab and some area of Haryana and Rajasthan. For Punjabi language many speech recognition systems exists but all have some kind of problems which have to solve out for better performance. This paper aims to discuss various other speech recognition methods that have been developed for isolated words, connected words and in other languages too. It also discusses the mechanisms implemented and the accuracy achieved respectively.

Keywords- *MFCC, HMM, Taksali Punjabi, HTK, ASR, LPC.*

I.INTRODUCTION

Speech is the vocalized form of human communication. To communicate with a machine we still require interfaces like keyboard, mouse and screen etc. operated with the help of software. Speech recognition technique has been explored widely for the past four decades as the degree of acceptance for such systems is high. Speech is considered to be one of the easiest and comfortable means of communication. It is an efficient way to recognize a person on the basis of speech. It is also an important biometric authentication process. To exchange any information between man and machine keyboards, pointing devices are required which is not convenient for a layman as it requires special skills. Hence, speech has provided a great platform to resolve this issue. Speech recognition is the process where the human speech is input into the system in analog form and the machine converts it into digital form to make it understandable. Many such systems have been developed and worked upon to maximize the ease of access for communication. Speech recognition system divided into two parts speaker dependent and speaker independent [4]. Speaker dependent recognize speech from one person and speaker independent system can recognize speech from any person. Another classification of speech recognition is speech to text [7,8], speech to signal and speech to speech. According to [8] every speech recognition system consists of two parts called as system training and testing. In training of system is trained for particular words which are tested in testing phase and accuracy mainly depends upon training of system. Feature extraction [10] is defined as the process to extract a small amount of data from a speaker's voice signal that can be used later to represent that speaker. It deals with retaining useful information of the signal while discarding redundant and unwanted information. Features such as power, pitch, and vocal tract configuration, the amplitude of the signal and energy of the frequencies are extracted. Feature matching involves the actual procedure to identify the

International Journal of Advance Research in Science and Engineering
Volume No.06, Special Issue No.(01), Nov 2017
www.ijarse.com

IJARSE
ISSN: 2319-8354

unknown speaker by comparing the extracted features from his/her voice input with the ones that are already stored in our database. In system testing[4] feature matching from coming speech signals with features saved in database. Feature matching means match feature vector which store in database against features of given input speech if it matches pre-defined function is perform else it request for training. MATLAB (matrix laboratory) is a fourth-generation high-level programming language and interactive environment for numerical computation, visualization and programming. MATLAB is developed by Math Works. It allows matrix manipulations; plotting of functions and data; implementation of algorithms; creation of user interfaces; interfacing with programs written in other languages, including C, C++, Java, and FORTRAN; analyze data; develop algorithms; and create models and applications. Mostly all systems are developed with MATLAB because it has building functionality to support user.

II.RELATED WORK

Ghai [1] analyzed efforts made by various researchers to develop automatic speech recognition systems for most of the Indo-Aryan languages and their applicability to Punjabi language is discussed so as to initiate a concrete work for Punjabi language 1500 isolated words were used for training mode as a knowledge base while performance was evaluated with 500 isolated Punjabi words. DTW Dynamic Time warping was used for recognition. Recognition accuracy was found to be 61% for isolated words and lesser for connected words. Dua[7] worked for Punjabi speech to text system for connected words on HTK 3.4.1 speech engine on the Linux platform. The overall performance was analyzed for both class and open environment. 6 unique speakers have been asked to record test data each consisting of 30-50 words. The performance of system recorded as Word Recognition Rate (WR) 95.8% and 95.4%, Word Accuracy Rate (WA) 94.1%and 91.6% and Word Error Rate (WER) 5.9% and 8.35% in class room and open environment respectively. Dua[8] worked for Punjabi automatic speech Recognition using HTK.The performance of the system was tested against speaker independent parameter by using two types of speakers: one who are involved in training and testing both and the other who are involved in only testing. The second parameter for checking system performance is in different environments. The system was tested in a class room and in open space. A total of 6 distinct speakers were used for this and each one was asked to speak 35-50 words. The results shown reveal that the implemented system performs well with different speakers and in different environments. The average performance of the system lies in the range of 94 % to 96% with word error rate4% to 6%. Kumar [3] worked for the recognition accuracy rate of the Punjabi interview speech corpus recognized the 461 sentences and 1227 Punjabi words. Out of 461 sentences, 6 sentences and out of 1227 Punjabi words, 14 words are failed while recognition. In this research authors used sphinx and java programming for GUI of recognition system. Ghai [2] developed HTK 3.4.1 speech engine used in implementing ASR systems. Two approaches of acoustic modeling were used: whole word models and tri-phone models. The word recognition accuracy of isolated word speech was 92.05% for acoustic whole word model based system and 97.14% for acoustic tri-phone model based system. The word recognition accuracy of connected word speech was 87.75% for acoustic whole word model based system and 91.62% for acoustic tri-phone model based system. The system used [5] a noise free room environment for data recording. MFCCs were used to extract features from the speech signal. System had been trained using HMM

which was considered to be one of the most efficient pattern recognition techniques. Viterbi Algorithm had been used for system testing as it was found to be the one that finds the most probable sequence of path. The authors proposed the system for recognition of 90 alphabets taken from Punjabi language with 9 speakers having Punjabi as their native language from the age group of 20-50 years. The accuracy of the system was evaluated in three different scenarios. The results were found to be 80%, 100% and 55% accurate for each scenario respectively. Kaur's [6] name entity recognition system in Punjabi language to seek and classify words which represent proper names in text into predefined categories like location, person name, organization, date, time, designation etc.

Table 1:-Accuracy of NER System

NE CLASS	P(%)	R(%)	F(%)
Person	74.52	62.86	65.67
Location	91.52	92.89	91.25
Organization	90.27	90.10	88.77
Designation	98.84	87.09	91.98
Date/Time	94.79	89.79	91.75
Total	89.98	84.55	85.88

The system was good results for location, organization, designation, date/ time NER's with average accuracy of 85%. The results for person name were not good as compared to other NEs because accuracy of that was around 62%.Shrawanker [10] compared the performances of MFCC and LPC features under VQ environment. Marathi speech database was recorded in noisy environment so that we compare MFCC and LPC in difficult conditions.

Table 2:-Accuracy of LPC Technique

Word	Speaker1	Speaker2
AAI	75%	73%
ANANAS	78%	74%
BAL	80%	78%
KSHATRIYA	81%	80%
AVERAGE	78.5%	76.25%

Table 3:-Accuracy of MFCC Technique

Word	Speaker1	Speaker2
AAI	98%	99%
ANANAS	100%	100%
BAL	100%	100%
KSHATRIYA	100%	100%
AVERAGE	99.5%	99.75%

Table 2 and 3 showed that recognition accuracy was more with MFCC. So MFCC can be thought of as better feature for Marathi language tutor application in speech recognition. Kumar's[15] spontaneous Punjabi speech model, the speech is basically non-planed and non-designed, there are generally depicted by repetitions, preservation, wrong start, half-spoken words and non-planned words, silence gap etc. In a system of Punjabi speech detection including vocabulary, the identification needs the evaluation among the audio signal of the utterance and the variety of utterances of the vocabulary.

Table 4:-Accuracy of spontaneous model

Spontaneous Punjabi speech	Total	Correct words /sentences	Errors	Recognition Accuracy
Punjabi words	691	630	61	91.17
Punjabi sentence	2115	1806	309	85.38

In study of "Automatic Speech Recognition System for Hindi Utterances with Regional Indian Accents". Features of the speech signal were extracted in the form of MFCC coefficients and Dynamic Time Warping (DTW) had been used as features matching techniques. The recognition results were tested for clean and noisy test data. Average accuracy for clean data was97.50 % while that for noisy data was 91.25 %.This paper showed that these approaches can be used to increase accuracy of noisy test data for automatic speech recognition system with regional Indian accents. The study in paper [17] shows the comparison between feature extraction and feature classification techniques as follows in tables.

Table 5:-Feature Extraction Techniques

S.No	Technique	Principal	Merits and De-merits
1.	Linear Predictive Coding	Model by all pole model	(a)Based on Basic Principle of Sound Production. (b)Performance degradation in presence of noise.
2.	Cepstral Coefficients	FFT Based	Not much Consistent with human hearing due to representation by linearly spaced filters.
3.	Linear Predictive Cepstral Coefficients	Modeled by pole Model	(a)Gives smoother Spectral Envelope and stable representation as compared to LPC. (b)Drawback due to linearly spaced frequency.
4.	Mel-Frequency Cepstral Coefficients	Filter Bank Coefficients	More information about lower frequencies due to Mel spaced filter banks hence behaves more like a human ear as compared to other techniques.

Table 6:-Feature Classification Techniques

S.No	Classifier	Type of Algorithm	Merits and De-merits
1.	Support Vector Machines	Supervised	(a)Beneficial in case of binary classification. (b)Poor Performance in speaker recognition due to its restriction to work with fixed length vectors.
2.	Hidden Makov Model	Unsupervised	(a)Computationally more complex and needs more storage space. (b)Needs more training data to deal with intersession issue.
3.	Vector Quantization	Unsupervised	(a)Memory requirement is feasible for real time Application. (b)Computationally less complex.
4.	Gaussian Mixture Model	Unsupervised	(a)Needs less training and testing data. (b)Compromise between DTW and HMM.

III. CONCLUSION AND FUTURE WORK

In conclusion, an efficient, accurate speech recognition system for regional languages like Punjabi is need of hour. The work purposed in this paper is step towards the development of such type of systems. The work may further be extended to large vocabulary size and with local words of Punjabi. Future work will include the development of continuous speech recognition, the development of faster training and testing and advanced user interface.

IV. ACKNOWLEDGMENT

REFERENCES

[1] Wiqas Ghai and Navdeep Singh, "*Analysis of Automatic Speech Recognition Systems for Indo-Aryan Languages: Punjabi a Case Study*", International Journal of Soft Computing and Engineering, *2*,2012, 379-385.

[2] Wiqas Ghai and Navdeep Singh, "*Phone Based Acoustic Modeling for Automatic Speech Recognition for Punjabi Language*", IJCA, *1 (3)*, 2013. 69-83.

[3] Yogesh Kumar and Navdeep Singh, "*Automatic Spontaneous Speech Recognition for Punjabi language Interview speech Corpus*",IJEME, *6*, 2016, 64- 73,.

[4] Harpreet Kaur, "*A Novel Speech Recognition system for Punjabi Language: A review*", International Journal of Computer Science and Information Technologies,*6(3),* 2015.

[5] Harpreet Kaur and Rekha Bhatia, "*Speech Recognition system for Punjabi language",* International Journal of Advanced Research in Computer Science and Software Engineering*, 5*, 2015.

[6]   Kamaldeep Kaur and Vishal Gupta," *Name Entity Recognition for Punjabi Language*", IJCA, *2*, 2012

[7] Mohit Dua and R.K Aggarwal,"*Punjabi Speech to Text System for Connected Words*", Fourth International Conference on Advances in Recent Technologies, 2015.

[8] Mohit Dua and R.K Aggarwal,"Punjabi *Automatic Speech Recognition Using HTK*", International Journal of Computer Science Issues, *9*, 2012.

[9] T.K Das and Khalid M.O Nahar,"*A Voice Identification System using Hidden Markov Model*", Indian Journal of Science and Technology**,** *9 (4)*, 2016

[10] Urmila Shrawankar and Dr.VilasThakare,"*Techniques for Feature Extraction in Speech Recognition System: A Comparative Study*", International Journal of Computer Applications in Engineering, Technology and Sciences, 2011

[11] JinalH.Tailor and DiptiB.Shah,"*Speech Recognition System Architecture for Gujarati Language*", International Journal of Computer Applications,*138*, 2016

[12] Ranu Dixit and NavdeepKaur,"*Speech Recognition Using Stochastic Approach: A Review*", International Journal of Innovative Research in Science, Engineering and Technology ,*2*, 2013

[13] Kumar Ravinder,"*Comparison of HMM and DTW for Isolated Word Recognition System of Punjabi Language*", springer, 2010, 244–252.

[14] Kanika Sharma and Parminder Singh," *Speech Recognition of Punjabi Numerals using Synergic HMM and DTW Approach*", Indian Journal of Science and Technology, *8 (27)*, 2015.

[15] Yogesh Kumar and Navdeep Singh," *An Automatic Spontaneous Live Speech Recognition System for Punjabi Language Corpus*", International Journal of Advanced Research in Computer Science and Software Engineering, *9 (20)*, 2016.

[16] Abhishek Thakur, Rajesh Kumar and Naveen Kumar,"*Automatic Speech Recognition System for Hindi Utterances with Regional Indian Accents: A Review"*, IJECT, *4*, 2013

[17] Gaganpreet Kaur, Dr.Dheerendra Singh and Gagandeep Kaur," *A Survey on Speech Recognition Algorithms*", International Journal of Advanced Research in Computer Engineering & Technology, *4*, 2015