

Review on Speech Recognition System for Indian Languages

Jinal H. Tailor
PhD Scholar
Sardar Patel University,
Vallabh Vidyanagar, India

Dipti B. Shah, PhD
G H Patel Post Graduate
Dept. of Computer Science & Technology,
Sardar Patel University,
Vallabh Vidyanagar, India

ABSTRACT

Speech recognition system is a natural way for the interaction of human to machine. Automatic Speech Recognition is advance way to operate computer without much efforts through speech only. In this paper survey related to indo Aryan languages usage for communicating directly with the machine has been performed. This mechanism includes various techniques and experimental results. Speech Recognition system is implemented for English, French, Spanish, German, Japanese and Chinese. Only little work has been performed for Indo-Indian languages like Gujarati, Marathi, Hindi, Tamil etc. Speech to Text is an emerging research area due to complexity and various frameworks of Indo-Aryan languages.

Keywords

Speech Recognition, Gujarati, Indo-Aryan language, Speaker Independent.

1. INTRODUCTION

Speech recognition system provides user friendly interface to the user. It will be more useful when it will be available in natural languages. Speech recognition system encourages use of technologies to the people who are not so comfortable in operating machines due to lack of knowledge and language barrier as well as to people with disabilities. Speech recognition in native languages provide convenient and hands-free environment to the user. Speech Recognition approach is used extensively to solve real-world challenges. So many factors affect accuracy and performance of speech recognition system. Due to different grammar rules, noisy environment, and pronunciations of speaker independent speech recognition system is a challenging job.

2. ANALYSIS OF RESEARCH WORK IN INDIAN LANGUAGES

In India there are 122 major languages and 1599 other languages spoken.

The survey has been done for the speech recognition system developed for the Indian languages. The summary of this survey is as below:

Patel J, Patel P and Virparia P (2013) They have developed speaker-independent system that accept telephony commands in Gujarati. They have used speaker independent 29 words for the experiment. Implementation is done using HMM based speech recognizer Sphinx4 toolkit. In the whole experiment 20 speakers are involved, 10 female and 10 male with the age criteria between 20 to 30 years. Total 29 words were used for testing as command. During the experiment, Average accuracy for female speakers is 83.79%. Average accuracy for

male speakers found 80%. The minimum accuracy of the system is 72.41% and after taking extra care system can achieve highest accuracy up to 96.55%. Average accuracy rate of all the words found 83.62%.

Patel H and Virparia P (2011) They have proposed a model to recognize small vocabulary speech in Gujarati language using English speech recognition engine. Aligned data is used to obtain initial acoustic model for the phones according to its framework. There are two kinds of models available, Acoustic model and language model. Bootstrapping technique is used to align Gujarati speech to English speech engine. In the model total 30 words used for recognition as small vocabulary. Total number of speaker 31 from which 12 female and 19 male. Overall recognition accuracy 88.71% from which average frequency of male is 90.88% and for female it is 85.28%. For the model age wise result is also considered and it is found near to average accuracy. The obtained result from the model is analyzed through various recognition experiments.

Patel H, Virparia P (2012) They have proposed model for automatic speaker identification in Gujarati. Model includes two major processes voice verification of speaker and identification. At registration phase voice model generated and stored on smart card that will be further used for verification. They have listed some verification errors which include stress, time varying, aging, sickness etc. model presents phases that include feature selection and measures and pattern matching. Two types of model are included stochastic and template. The proposed model implemented using MARF (Modular Audio Recognition Framework) which includes collection of algorithms of sound, speech and natural language processing.

Sahani S, Ramamohan S, Pradhan V.: They have implemented a hybrid system for recognizing and generating Gujarati numeric with using neural network and Java API. In the system as an input Gujarati handwritten digit is entered in the form of matrix of 5X5. The system network receives 25 Boolean values and 25 element of input vector. From the output it receives 10 elements vector each represent Gujarati numeric. With the use of neural network backpropagation network method is used to recognize Gujarati numeric. With using this system they have concluded that recognizing Gujarati digits using ANN back propagation method enhance in performance when using with 2 hidden layers. With concern to accuracy MLP method gives best accuracy when used with 2 hidden layers. In the system no. of layers is proportional to no. of epochs. With using Java API entered Gujarati numeric can be converted into sound.

Pandit P, Bhatt S. (2014) They have developed Automatic Speech Recognition for Gujarati numeric with using Time Warping. For this system MFCC and DTW techniques are used for recognition. Recording of 10 different speakers were done using Audacity, open source software. Sampling rate of recording was 16000 HZ with using the mono channel. Recorded speech signals cut at zero-crossing and as output it will give 100 speech signals. Feature extraction is done with using MFCC technique. The output of the technique is received in the form of cepstral coefficient that is used in DTW method. They have concluded that 8 out of 10 numbers digits by the selected speaker are matching with the template. Success ratio of the system is 84.44%. it can be improve up to 95.56%.

Ghai W, Singh N (2013) They concluded that there is a long way to go so as to enhance the performance standards set

for other Indo-Aryan languages. They observed that use of techniques Cooperative Heterogeneous ANN Architecture, Maximum Likelihood Linear Regression, Extended MFCC and Learning Vector quantization are helping the researchers to get improved recognition performance of speech recognition systems. The result analyzed Provide Good accuracy level.

Akila, E. Chandra (2013) They have described about some speech synthesis and speech recognition algorithms and compares their performance based on accuracy and quality. In speech recognition DTW and HMM algorithms are compared with respect to accuracy. Comparative study of CELP and MBROLA algorithm of speech synthesis based on quality is also done. As a result accuracy found in DTW is 77% and in

HMM 92%. Comparing quality MBROLA algorithm provides good quality at all bit rate.

Mishra A, Chandra M, Biswas A, Sharan S.: They have developed speaker independent Hindi digits recognizer system using Revised perceptual linear prediction (RPLP), Bark frequency cepstral coefficients (BFCC) and Mel frequency perceptual linear prediction (MF-PLP). The recognizing performance compared with MFCC and PLP features. MFCC features are calculated by using feature extraction tool of Hidden Markov Model Toolkit (HTK). It is concluded that MF-PLP gives best recognizing performance compare to other feature extraction method because it incorporates Mel filter into a perceptual linear feature extraction method in clean as well as for noisy database.

Sivaram G, Samudravijaya K: They have studied the impact of online speaker adaptation on the performance of a speaker independent, continuous speech recognition system for Hindi language. The speaker recognition is executed using the Maximum Likelihood Linear Regression (MLLR) transformation approach. The MLLR transform based speaker adaptation technique is found to significantly improve the accuracy of the Hindi ASR system by 3%. After the experiment they have concluded that MLLR transform based speaker adaption of Hindi speech models indeed decreases the recognition error by a factor of 0.19

3. ANALYSIS OF RELATED WORK IN INDIAN LANGUAGES

<u>Authors</u>	<u>Year</u>	<u>Spoken Language</u>	<u>Methods/ Techniques</u>	<u>Category of Tokens</u>	<u>Accuracy</u>
Sarma and Sarma	2011	Assamese	LVQ learning vector quantization	Speech	95%
Sarma et. Al	2008	Assamese	ANN, RNN	Speech	RNN – 96% ANN – 97.5%
Chowdhury	2010	Bengali	CMU Sphinx Tool	Speech	90.65% By Varying Speaker - 86.79%
Hasnat et. Al	2007	Bengali	MFCC, HTK Tool (HMM)	Speech	For Isolated Speech 90% For Continuous Speech 80%
Samudravijaya	2000	Hindi	HMM	Words, Sentence	Word Accuracy 70% Sentence Accuracy 50%
Kumar and aggarwal	2011	Hindi	HMM	Words	Word accuracy 94.63 %
Aggarwal and dave	2011	Hindi	MFCC, HLDA hetroscedastic linear discriminant	Speech	Word Model – 74.6% Phoneme Model – 76% Triphone Model – 79.2%
Gawali et.al	2010	Marathi	MFCC, DTW	Speech	MFCC – 94.65% DTW – 73.25%
Gaikwad et.al	2011	Marathi	IVR	Sentences	88%

Mohanty and Swain	2011	Oriva	MFCC,HMM	Words	Seen data 94.72% Unseen data 78.23 %
Nadungodage and Weerasinghe	2011	Sinhala	HTK	Word Sentence	Word accuracy – 96 % Sentence accuracy – 75 %
Raza et.al	2009	Urdu	Sphinx 3 trainer and decoder	Speech	58.4%
Sarfaraz et.al	2010	Urdu	CMU Sphinx open source toolkit	Speech	Male – 39.8% Female – 54.2%
Kumar	2004	Punjabi	vector quantization and dynamic wrapping, LPC analysis	Word	Recognition accuracy for isolated words – 61 %

Despite of advanced technology of speech recognition system it is not implemented in Gujarati Indo- Aryan language which is spoken by 54.6 million people in India. According to CIA (Central Intelligence Agency) Gujarati Language speakers in India are above 4.5 % of population. Gujarati, being a widely spoken language is still trailing in research and development in the field of Speech Recognition. Apart from limitations and complexity of Gujarati Framework, implementation of speech recognition system has wide area of research.

4. CONCLUSION

As speech recognition system is efficient use of technology, but many people are not able to take benefit of it because of language barriers, disabilities as well as lack of operating knowledge of the machine. Hands free environment with the native language facility will provide utilization of technology to various fields of user. An implementation challenge that occurs with native language speech recognition is complex framework and huge vocabulary. Speech recognition system in indo-Aryan languages includes complex grammar and pronunciation identification challenges. The complexity also includes digits and symbol representation. Some limitations involve such as to accuracy and acceptance. By finding proper noise removal techniques accuracy level can be increased. Hence, there is a scope for developing the Speech Recognition System for Gujarati Language.

5. REFERENCES

- [1] Aggarwal, R., and Dave M., —"Using Gaussian Mixtures for Hindi Speech Recognition System", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 2, No. 4, December 2011.
- [2] Akila A.,E. Chandra. , - "Isolated Tamil Word Speech Recognition System Using HTK", International Journal of Computer Science Research and Application, Vol. 3, Issue 2,Pages 30-38
- [3] Chowdhury, S., —"Implementation of Speech Recognition System for Bangla", BRAC University, DHAKA, Bangladesh, August 2010.
- [4] Gaikwad S., Gawali, B., Mehrotra S., —"POLLY CLINIC INQUIRY SYSTEM USING IVR IN MARATHI LANGUAGE", International Journal of Machine Intelligence, ISSN: 0975-2927 & E-ISSN: 0975-9166, Volume 3, Issue 3, 2011, pp-142-145.
- [5] Gawali, B., Gaikwad S., Yannawar P., Mehrotra S., —"Marathi Isolated Word Recognition System using MFCC and DTW Features", International Conference on Advances in Computer Science 2010, DOI: 02.ACS.2010. 01.73.
- [6] Ghai W., Singh N., - "Analysis of Automatic Speech Recognition Systems for Indo-Aryan Languages : Punjabi A Case Study", International Journal of Soft Computing and Engineering, Vol.2, Issue 1, March 2012
- [7] Hasnat, M., Molwa, J.,Khan, M., —"Isolated and Continuous Bangla Speech Recognition: Implementation, Performance and application perspective",2007.
- [8] Kumar, K.,Aggarwal R., —"Hindi Speech Recognition System Using HTK", International Journal of Computing and Business Research, ISSN (Online): 2229-6166, Volume 2 Issue 2, May 2011.
- [9] Kumar R., Singh C., Kaushik S., —"Isolated and Connected Word Recognition for Punjabi Language using Acoustic Template Matching Techniquel, 2004.
- [10] Kumar, R., —"Comparison of HMM and DTW for Isolated Word Recognition System for Punjabi Language", International Journal of Soft Computing 5(3):88-92, 2010.
- [11] 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
- [12] Mohanty S., Swain B., —"Continuous Oriya Digit Recognition using Bakis Model of HMM", International Journal of Computer Information Systems, Vol. 2, No. 1, 2011.
- [13] Mohanty S., Swain B., —"Markov Model Based Oriya Isolated Speech Recognizer",An Emerging Solution for Visually Impaired Students in School and Public Examinationl, Special Issue of IJCCT Vol. 2 Issue 2, 3, 4; International Conference On Communication Technology-2010.

- [14] Nadungodage T., Weerasinghe, R., —Continuous Sinhala Speech Recognizer, Conference on Human Language Technology for Development, Alexandria, Egypt, May 2011.
- [15] Pandit P., Bhatt S., - “Automatic Speech Recognition of Gujarati Digits using Dynamic Time Warping”, International Journal of Engineering and Innovative Technology, Vol. 3, Issue 12, June 2014
- [16] Patel H., Virparia P., - “A Small Vocabulary Speech Recognition for Gujarati”, International Journal of Advanced Research in Computer Science, Vol. 2, No.1, Jan-Feb 2011
- [17] Patel H., Virparia P., - “Generic Model for Text Dependent Automatic Gujarati Speaker Recognition”, International Journal of Emerging Trends & Technology in Computer Science, Vol. 1, Issue 3, September – October 2012
- [18] Patel J., Patel P., and Virparia P., - “Voice Enabled Telephony Commands using Gujarati Speech Recognition”, International Journal of Advanced Research in Computer Science and Software Engineering”, Vol. 3, Issue 10, October 2013
- [19] Purnima P., Bhatt S., - “Automatic Speech Recognition of Gujarati Digits using Dynamic Time Warping”, International Journal of Engineering and Innovative Technology, Vol. 3, Issue 12, June 2014
- [20] Sahani S., Ramamohan S., and Pradhan V., - “On Speech Synthesis for Gujarati Numeric”, Bulletin of the Marathwada Mathematical Society, Vol. 13, No.2, pages 53-60, December – 2013
- [21] Sarfraz H., Hussain S., Bokhari R., Raza A, Ullah I., Sarfraz Z., Pervez S., Mustafa A., Javed I., Parveen R., —”Large Vocabulary Continuous Speech Recognition for Urdu”, International Conference on Frontiers of Information Technology, Islamabad, 2010.
- [22] Samudravijaya, K., —”Computer Recognition of Spoken Hindi”, Proceeding of International Conference of Speech, Music and Allied Signal Processing, Triruvananthapuram, pages 8-13, 2000.
- [23] Sarma M, Sarm K., —”Assamese Numeral Speech Recognition using Multiple Features and Cooperative LVQ – Architectures”, International Journal of Electrical and Electronics 5:1, 2011.
- [24] Sarma, M., Dutta, K., Sarma, K., —”Assamese Numeral Corpus for Speech Recognition using Cooperative ANN Architecture”, International Journal of Electrical and Electronics Engineering 3:8 2009.
- [25] Sivaraman G., Samudravijaya K., - “Hindi Speech Recognition and Online Speaker Adaption”, International Conference on Technology System and Management, ICTSM-2011, IJCA.
- [26] Raza, A., Hussain, S., Sarfraz, H., Ullah, I., and Sarfraz, Z., —”An ASR System for Spontaneous Urdu Speech”, Proceedings of O-COCOSDA’09 and IEEE Xplore, 2009.