ISSN:0975-9646

A Novel Speech Recognition system for Punjabi language: A review

Harpreet Kaur
Punjabi University Regional centre for Information technology and management, Mohali

Abstract - Punjabi being one of the most widely used languages in media and communication, a speech recognition system for this language is need of the hour. Speech recognition is a biometric authentication process where voice characteristics are used as the attribute. Automatic speech recognition is a process where speech signals are automatically converted into the corresponding sequence of words in text or the converted signal is used to operate machine with commands. 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: Automatic speech recognition, Punjabi language, HMM, HTK, DTW, MFCC.

I. INTRODUCTION

Speech being the easiest mode of communication is considered to be an efficient way to identify a person on the basis of speech alone. During the coming years, it is hoped that speech recognition will make it possible to verify the identity of a person accessing a system. It will allow automated control of services by voice, such as banking transactions and also control the flow of private and confidential data in industries. It is one of the most useful biometric authentication processes where the characteristics of the human voice are used as attributes [9].

Speech recognition systems can further be classified as speaker-dependent or speaker-independent system [10]. A speaker-dependent system recognizes speech from one particular speaker's voice whereas a speaker-independent system recognizes speech from anybody. Speech recognition systems can be implemented for an isolated word or for continuous speech.

The process of speech recognition involves two phases, a training phase and a testing phase [8]. Training involves teaching the system, where a dictionary is build and an acoustic model is used to recognize each word that the system processes. During the testing stage, an acoustic model is used to recognize isolated words using a classification algorithm.

Speech recognition mainly deals with identifying the speaker which is done on the basis of the individual information included in the speech waves. This process involves two basic modules, feature extraction, and feature matching.

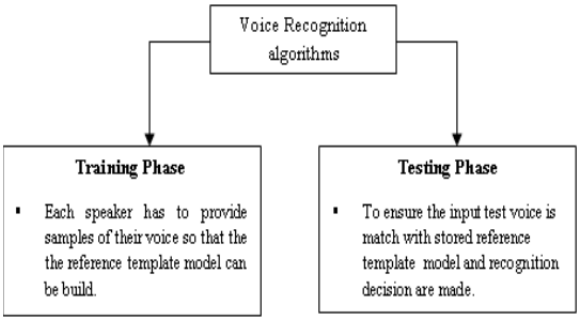

Fig.1 Voice Recognition Algorithms [8]

Feature extraction 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 [9]. Features such as power, pitch, vocal tract configuration, the amplitude of the signal and energy of the frequencies are extracted [5][6]. Feature matching involves the actual procedure to identify the unknown speaker by comparing the extracted features from his/her voice input with the ones that are already stored in our database [9]. Feature extraction is done using various methods, namely [11]:

1. Mel Frequency Cepstrum Coefficients (MFCC)
2. Linear Predictive analysis (LPC)
3. Relative spectra filtering of log domain coefficients (RASTA)
4. Perceptual linear predictive coefficients (PLP)
5. Mel scale Cepstral analysis (MEL)
6. First order derivative (DELTA)
7. Power spectral analysis (FFT) etc.

The various techniques used for speech recognition are Hidden Markov Model (HMM), Dynamic Time Warping (DTW), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM) and Decision Trees [11].

II. RELATED WORK

The authors of [1] developed a real-time, isolated word Automatic Speech Recognition system (ASR) to work in both speaker-dependent and speaker-independent environments for an Indian regional language, Punjabi. A statistical approach using the HTK 3.4 toolkit based on Hidden Markov Model (HMM) installed on Linux environment Ubuntu 11.10. The presented work used JAVA platform for the implementation of test module so as

to make the system more interactive and fast. The system was trained for 115 distinct Punjabi words and the data was collected from eight speakers and was tested by using samples from six speakers. 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 while the other ones who are involved in testing only. The system was tested in a class room and in open space. The overall system performance was evaluated using the HResult tool by HTK where the Word Error Rate (WER) was calculated and the results obtained were 95.63% and 94.08% accurate with a WER of 4% to 6%.

In [2] authors worked for automatic speech recognizer for continuous speech sentences and used Tri-Phone based acoustic modeling on HTK 3.4.1 speech engine on the Linux platform. The sequence of three words, L-X+R is the Triphones. It defines the left context (L) and right context (R) of the current word (X). The overall performance was analyzed for both seen and unseen data. 100 sentences of Punjabi language were used. The performance was evaluated in three phases, seen data in the first two phases while unseen data in the third one. The results so obtained for overall sentence recognition accuracy and word recognition accuracy were 82.18% and 94.32%.

In [3] authors 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.

The authors of [5] developed a connected words speech recognition system for Hindi language based on MFCC using HTK v3.4. It was developed in the Ubuntu 10.04 operating environment which is a Linux platform. HTK was used to develop the system that uses HMMs for recognition. The system was trained using a vocabulary size of 102 words and training data was collected from 12 different speakers. Testing data was prepared separately by a set of five speakers. System performance was evaluated by testing in room environment, open space, lab room, class room environment as well as market environments. The overall word accuracy and word-error rate of the system are 87.01% and12.99% respectively. Word correction rate was found to be 90.93%.

The authors of [8] developed a speaker-dependent continuous speech recognizer for a non-Latin language, Sinhala. The data was recorded with a single female voice consisting of 983 distinct continuous Sinhala sentences which were used for training. The vocabulary size of the data is 4055 words. Another 106 words were generated using the words in the data set which were used for training and evaluation purpose. The results so obtained were 75.74% and 96.14% for sentences and words respectively.

In [12] authors describe an automatic speech recognizer for isolated word speech and connected word speech using a triphone based acoustic model and compares its performance with acoustic whole word-based model based automatic speech recognition system. The vocabulary comprises of 200 words for both training and testing experiments, selected in a phonetically balanced way so that correct pronunciations can be covered for all the phones of Punjabi language. To create word level transcriptions, a Perl script was used in the form of an MLF (Master label file). For the training data samples 3 male and 3 female speakers were included and 5 utterances of each word from each speaker were recorded, a total of 6000 speech files were recorded. The performance was evaluated in 2 phases for isolated word and connected word speech recognition. Percentage accuracy was found to be 92.05% for acoustic word model and 97.14% for acoustic triphone model for isolated word speech recognition. Percentage accuracy for connected word speech recognition was 87.75% for acoustic word model and 91.62% for acoustic triphone model respectively.

III. PERFORMANCE PARAMETERS

To analyze the system performance, HResult tool is provided by HTK. It compares the accuracy of the system. The transcription file of the HVite tool, which is used to generate the output is compared with the corresponding original reference transcription file.

The performance of a speech system is calculated as follows:

$$\textit{Percentage correct (PC)} = \frac{N-D-S}{N} \times 100 = \frac{H}{N} \times 100$$

where N is the number of words in the test set, D is the number of deletions, S is the number of substitutions and H is the number of correct labels.

PC gives the percentage of the correctly recognized word.

$$\textit{Percentage accuracy (PA)} = \frac{N-D-S-I}{N} \times 100 = \frac{H-I}{N} \times 100$$

where I is the number of insertions.

PA gives the word accuracy rate.

$$\textit{Word error rate (WER)} = \frac{S+I+D}{N} \times 100$$

$$= 100 - \text{Percentage accuracy}$$

IV. CONCLUSION

Speech recognition has gained popularity in the past few decades and the demand for ASR (automatic speech recognition system) has greatly increased. A Lot of research is being carried out for various languages. Punjabi being a widely spoken language requires a lot of research. So far work has been done for isolated words, connected words and continuous speech, however, an efficient system for continuous speech has to be developed yet. This paper discusses various speech recognition systems for different languages. It suggests a few methods in different environments and their results have been discussed. The major requirement is to develop an ASR for connected speech with more accurate recognition rate.

REFERENCES

[1] Dua, M., Aggarwal, R. K., Kadyan, V., & Dua, S., "Punjabi Automatic Speech Recognition Using HTK", *International Journal of Computer Science Issues,* vol. 9, issue 4, no. 1, pp.359–364, 2012.

[2] Ghai, W. & Singh, N., "Continuous Speech Recognition for Punjabi Language", *Internal journal for computer applications(0975-8887),* vol. 72, no. 14, pp. 23–28, May 2013.

[3] 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 (IJSCE)*, vol. 2, no.1, pp. 379–385, March 2012.

[4] Gupta, R., "Speech Recognition for Hindi", M.Tech. thesis, Dept. Comp. Sci. Eng., Indian Institute of Technology, Bombay, 2006.

[5] Kumar, K., Aggarwal, R. K., & Jain, A., "A Hindi speech recognition system for connected words using HTK", *International Journal of Computational Systems Engineering*, vol. 1, issue 1, pp.-25. doi:10.1504/IJCSYSE.2012.044740, 2012

[6] Muskan & Aggarwal, N., 2014, "Punjabi Speech Recognition : A Survey", *Proceedings of the International Conference on Advances in Engineering and Technology,* vol. 3.

[7] Muda, L., Begam, M., & Elamvazuthi, I., " Voice Recognition Algorithms using Mel-Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", *Journal of computing,* vol. 2, no. 3, pp. 138–143, 2014.

[8] Nadungodage, T., & Weerasinghe, R., "Continuous Sinhala Speech Recognizer", *Conference on Human Language,* Retrieved from http://hltd.org/pdf/HLTD201123.pdf , pp. 2-5, May 2011.

[9] Rath, P. G. S., "Real-time speaker recognition using MFCC and VQ", M.Tech. thesis, Dept. Elec. Eng., National Institute of Technology, Rourkela, 2008.

[10] Saini, P., & Kaur, P., " Automatic Speech Recognition: A Review", *International Journal of Engineering Trends and Technology*, vol. 4, no. 3, pp.132–136, 2013.

[11] Singh, B., Kapur, N., & Kaur, P., " Speech Recognition with Hidden Markov Model : A Review", *International journal of advanced research in computer science and software engineering*, vol. 2, issue 3, 2012.

[12] Ghai, W. & Singh, N., "Phone based acoustic modeling for automatic speech recognition for Punjabi language", *Journal of speech sciences,* vol. 1, no. 3, pp. 69–83, 2013.

[13] Vimala, C., & Radha, V., "Suitable Feature Extraction and Speech Recognition Technique for Isolated Tamil Spoken Words", *International journal of computer science and information technologies,* vol. 5, no. 1, pp 378–383, 2014.