International Journal of Artificial Intelligence and Neural Networks – IJAINN
Volume 4 : Issue 3 *[ISSN 2250 – 3749]*
Publication Date : 30 September, 2014

Punjabi Speech Recognition: A Survey

By Muskan and Dr. Naveen Aggarwal

Abstract – **As Punjabi language is one of the most widely used languages in media and communication, its speech recognition is need of the hour. Hence, survey has been carried out for Punjabi speech recognition. In this, work has been carried out from boundary detection of isolated word recognition from Historical Perspective to that of present scenario. However, this has been limited to constraints and assumptions. This paper discusses the related work and future challenges.**

Keywords – Automatic Speech Recognition, Punjabi language, Hidden Markov Model, Dynamic Time Wrapping, MFCC, LPC.

I. INTRODUCTION

Speech is the source of communication. It carries data in the form of signals. Automatic Speech Recognition [1] (ASR) is the automatic speech transformation system which translates the audio input into text form. Identifying the words being spoken, mapping text to speech, verifying or identifying the speaker etc. are some of the activities being covered in speech processing. Although great amount of work has been performed in English language for speech recognition, Punjabi, being one of the most widely used languages, deserve to seek attention of researchers in terms of speech processing. Speech Recognition has two phases namely: Preprocessing and Post-processing which are discussed further. Accuracy, noise removal, information retrieval and varying bit rate are some of the most considerable part of speech recognition challenges.

1.1 TERMINOLOGY

The basic terminology is enlisted before proceeding with Punjabi Speech Recognition detailed study. **Phoneme** [8] is the basic unit of speech. During pronunciation, effect of two consecutive phonemes on each other is called **Co-articulation effect**. A **syllable** is the bigger unit which contains combination of phonemes. **Morphemes and Words** are more large units which contains combination of syllables. **Speech Corpus** is the database or collection of audio files. **Dictionary** contains the words being spoken. There are two types of speeches namely **speaker dependent** and **speaker independent**. There are two types of audio files: **training data** based on which the models are trained and **testing data** based on which we test the performance of the model. Usually speech models are trained before performing testing.

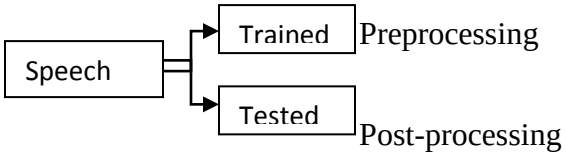
1.2 PUNJABI – ASR

Punjabi language is popular Indo-Aryan tonal Language. It has syllabic-centered nature. It is mostly spoken in Punjab and other republic parts of India and Pakistan. It is written from left to write. This language has 32 different dialects. Being tonal, differentiation in emotional emphasis in

pronunciation changes the meaning of speech. Gemination of consonants takes place in order to lay elongated and lengthy effect. Its phoneme inventory contains 10 vowels, 25 consonants, 7 diphthongs and three tones (deep-rising, high-falling and mid tone). Diacritics are used in combination with consonants in order to produce corresponding sound. The initial area of interest corresponds to reliable recognition due to boundary detection. This is usually done with the help of Syllables. There are seven types of syllables: V, VC, CV, VCC, CVC, CVCC, CCVC in Punjabi. Here V corresponds to vowel which is nucleus part and C corresponds to consonant. Left part of vowel V is onset and right part is coda. There are three types of sounds obtained in every audio file namely voiced, unvoiced and silence. Voiced contains relevant data or information, unvoiced sound contains noise or irrelevant sound and silence is integral part of speech signal which separates voiced and unvoiced part. These sounds are differentiated on the basis of energy level and thus, boundaries are detected. Thereafter work is further carried out to improve the accuracy of isolated word recognition in Punjabi language. MATLAB (Voicebox), Sphinx and HTK are some of the most widely used tools for speech recognition.

II. WORKFLOW

Speech Recognition undergoes various steps. There are usually two phases namely preprocessing and post processing. In preprocessing phase, speech is given as input signal. This signal is mapped in digital form i.e. sampled and quantized. Sampling frequency of speech varies from one format to another. After capturing audio signal, framing is performed followed by windowing. After this, feature extraction techniques are applied based on which features are being extracted. Usually, for Punjabi speech, MFCC and LPC features are being extracted. With this, different models are being trained. Recently work is being carried out in HMM, DTW and Neural Networks. After being trained, the models are set as persistent data. Also, we map the speech signals onto dictionary in order to obtain its grammar.



However, in post- processing phase, speech signal is given as testing data. In this, we extract feature vectors and further map it with the trained model. Further, we calculate distance between the testing file feature vectors and most resembling word vectors found in modeled file. This is done by using dictionary and grammar framed during preprocessing phase. If the distance is more than threshold, it is rejected else it is accepted. Finally, performance is calculated on the basis of percentage of correct observations as follows:

$$WER=(S+D+I)/N$$

Where S means number of substitutions; D indicates number of deletions; I identifies number of insertions and N corresponds to number of words of reference.

III. EVOLUTION OF DIFFERENT APPROACHES

A brief introduction of speech recognition work has been described. Initially, Speech

Recognition was introduced at Bell Laboratories by Davis et. al [11] in 1952. This work proposed the circuit for isolated work recognition system. This used to record spoken digits with the help of telephonic system.

3.1 HISTORICAL PERSPECTIVE

In 1971, Atal et. al [12] introduced linear combination of coefficients in order to extract features which work well for high pitched voices. Further Coker [13] improved linear prediction residual feature extraction technique for isolated word recognition. It removed redundancy from storage and thereby improved accuracy.

Rabiner et. al [14] worked on differentiating voiced, unvoiced and silence sounds from each other using LPC feature vectors. Later in 1974, they proposed [15] new algorithm for detecting endpoints on the basis of Zero crossings and energy. In 1981, Furui [16] proposed a new combination of LPC feature vector extraction along with DTW technique in order to recognize speech. In 1984, Bush et. al [17] proposed techniques to segment the speech automatically and manually with 96% to 97% of accuracy.

In 1978, H. Sakoe [28] proposed new dynamically programming algorithm which has been proved to be better than previous ones till date. In 1986, Rabiner et. al [18] explained the evolution of HMM which was used in 1960's and also introduce how it is being used for speech recognition. In 2000, Harb [19] introduces the neural network implementation for speech recognition system. Neural network was found to be better than that of HMM. In 2005, Axelrod et. al [20] introduced Gaussian Mixture Model which is state distribution of HMM.

In 2012, Mohamed et. al [21] proposed deep neural networks, an added advantage over GMM.

3.2 INDIAN LANGUAGES

In 1996, Raman Rao et. al [22] proposed new techniques for detecting word boundary. For Indian languages, accuracy was found to be 85%. In 2005, Patil et al [23] developed considerable speech corpora for Indian languages. In 2010, Ranjan [24] proposed DWT after applying LPC for feature extraction. He calculated codebooks by vector quantization. This work was performed for Hindi isolated word recognition.

In 2012, Bhardwaj [25] used MFCC and k-clustering algorithm for preparing codebooks of pre-trained data. Thereby, Viterbi algorithm was used for speech recognition in HMM. In 2000, Pruthi et al. [26] proposed 'swaranjali', the Hindi isolated word recognition system which used LPC and HMM for recognition. Also, during training, they used VQ for preparing codebooks. In 2013, Tripathi et al [27] worked for Hindi speech recognition system by extracting features using LPC and MFCC. She recognized the speech by using HMM and used HTK tool to implement the same.

3.3 PUNJABI LANGUAGE

In Punjabi Automatic Speech Recognition, the work has been initialized with segmentation i.e. boundary detection. In 2010, Path breaking work in Punjabi – ASR has been initialized by Kumar [1] who compared DTW and HMM for Isolated

Punjabi Word Recognition. He used MFCC and LPC feature extraction techniques for 300 samples over 8 KHz sampling rate. He found that HMM performs poorer recognition accuracy than DTW with 91.3% and 04.0% accuracy respectively over 500 samples. However, he stated that accuracy for HMM can be improved by increasing size of codebook. In July 2012, Dua et al. [2] built isolated word automatic speech recognition system using HTK toolkit for training phase. He used Java for betterment of user interface and hence, for testing the data. The system was developed in real time scenarios and is word based. The average performance achieved by the system lies in between 94% and 96%. In May 2013, Ghai et. al [3] developed a new model using HTK 3.4.1 for continuous speech recognition. They used tri-phone based acoustic modeling approach to obtain 82.18% and 94.32% accuracy at sentence level and word level respectively. In June 2013, Sharma et. al [4] describes the automatic continuous speech segmentation of Punjabi speech into syllables using the negative derivative of Fourier transformation i.e. group delay function. This is done after analyzing the characteristics of speech based on zero crossing rate and short term energy. In this the information in speech signal has been represented by Fourier analysis. In July 2013, Kaur et. al [5] developed a system in MATLAB for segmentation of syllables in speech signals. After identifying the start and end of the syllable, this system automatically labels the syllable without encountering any manual segmentation or labeled speech corpus. This is based on calculation of short time energy function for sampled frames. The automatic labeling is mapped with manual labeling with minor errors.

IV. COMPARISON

Work	Author	Input Dataset	Algorithm/ Tools Used	Efficiency of Techniques	Type of Natural Language
Isolated Word Recognition	K. H Davis et al [11]	Telephone digits	Frequency bands (Energy based)	97% to 99%	English
	L R Rabiner et. al [29]	Isolated Word	LPC and Improved DTW	Better than DTW	English
	Tarun et. al [26]	vocabulary of Hindi digits	Noise Elimination, LPC, HMM, VQ codebook	84.27%	Hindi
	Dua et. al [2]	Isolated Word in real environment	MFCC and HMM used. Java and HTK used	94.08% to 95.63%	Punjabi
Syllable Segment-ation	Bush et. al [17]	Isolated Digits	VQ using codebook sequences	96% to 97%	English
	Rabiner et. al [14]	Speaker independent words	LPC and Energy distance STE and ZCR, group	95%	English

	Sharma et. al [4]	Isolated Punjabi Speech: Words	delay	(Review)	Punjabi
Endpoint Detection	Rabiner et. al [15]	54 word vocabulary. Isolated words	Zero Crossing Rate and Energy	Good Accuracy	English
	G V Rao [22]	Isolated word as input	Pitch frequency variation	Indian: 85%	Indian
	Kaur et. al [5]	Isolated Word	STE and Endpoint detection algorithm	Accurate, Comparable	Punjabi
Speech Recognition LPC/ DTW	S. Furui [16]	Telephone Speech	LPC and DTW	99.33% and above	English
Speech Recognition MFCC/ HMM	Rabiner et. al [18]	Speech	HMM	(introduced)	English
	Ishan [25]	Hindi isolated words	MFCC and HMM	97.5%- 99%	Hindi
	Ghai et. al [3]	Continuous Punjabi Speech	MFCC, HMM, VQ, Acoustic Template Matching	82.18%, 94.32%	Punjabi
Compare: HMM / DTW & MFCC/ LPC	S. C. Sajjan et. al [9]	Limited Vocabulary isolated word	LPC, MFCC, DTW, HMM	LPC: 91% MFCC: 94%	English
	R. Kumar [1]	500 Punjabi isolated word vocabulary	LPC, MFCC, DTW, HMM	HMM: 92% DTW: 96%	Punjabi

Table 1: Comparison of different type of work carried out in different languages.

As speech recognition undergoes two phases, different feature extraction techniques including Mel Frequency Cepstral Coefficient (MFCC) and Linear Predictive Coding (LPC) played a pivotal role in analyzing and drawing good results. These techniques undergo framing, windowing and various Fourier transformations. Being an audio signal, speech recognition is recognized on the basis of different models mapped during training. These models include Hidden Markov Model, dynamic time wrapping technique, vector quantization etc. Short Term Energy and zero crossing rate were used to determine the syllable boundaries. All these techniques have been compared among different languages in Table 1. Also, different research papers have been reviewed in order to understand how different research works have been proposed in Punjabi language.

V. DISCUSSION

In Punjabi ASR concept, there are various challenging areas apart from those which have already being covered by researchers. These domains include variable sampling rate, gender identification, continuous speech recognition with nonlinear

alignment, noise removal, identifying co-articulation effect, speech synthesis. Also, for Punjabi Speakers, there are various applications in which this area is in demand for instance, forensic laboratories which identify or verify the target, mapping of Punjabi video into text format, retrieving information from Punjabi audio or video, converting text into Punjabi Speech. For actual recognition, there is need of Punjabi dictionary and Punjabi grammar which have not been prepared yet. However, accuracy is the most concerned area of research.

REFERENCES

[1] R. Kumar, "Comparison of HMM and DTW for isolated word recognition system for Punjabi language", *Proceedings of IJSC*, vol. 5, no. 3, pp.88 - 92, 2010.

[2] M. Dua, R. K. Aggarwal, V. Kadyan and S. Dua, "Punjabi automatic speech recognition using HTK", *Proceedings of IJCSI*, vol. 9, no. 1, July 2012.

[3] W. Ghai and N. Singh, "Continuous speech recognition for Punjabi language", *Proceedings of IJCA*, vol. 72, no. 14, May 2013.

[4] A. Sharma and A. Kaur, "Automatic segmentation of Punjabi speech into syllable-like units using group delay A Review", *Proceedings of IJCSET*, vol. 4, no. 6, ISSN: 2229-3345, June 2013.

[5] G. Kaur, P. Singh and A. Kaur, "Syllable boundary detection system for Punjabi language", *Proceedings of IJARC*, vol. 1, no. 2, July 2013.

[6] G. Kaur and P. Singh, "A technique to detect syllable boundary in a wave file", *Proceedings of IJCSCE special issue on NCRAET- 2013*, ISSN 2319 – 7080.

[7] L. Rabiner and R. Schafer, "Introduction to digital speech processing", *Foundations and Trends in Signal Processing*, Journal of ACM vol. 1, no. 1-2, pp. 1–194, 2007.

[8] L. Rabiner and B. H. Jaung, *Fundamentals of Speech Recognition*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[9] S. C. Sajjan and C. Vijaya, "Comparison of DTW and HMM for isolated word recognition", *Proceedings of International Conference on Pattern Recognition Informatics and Medical Engineering (PRIME)*, IEEE, pp. 466-470, 2012.

[10] W. Ghai and N. Singh, "Analysis of Automatic Speech Recognition systems for Indo-Aryan languages: Punjabi - A case study", *Proceedings of IJSCE*, vol. 2, no. 1, March 2012.

[11] K. H. Davis, R. Biddulph and S. Balashek, "Automatic recognition of spoken digits," *J.A.S.A.*, vol. 24, no. 6, pp. 637-642, 1952.

[12] B. S. Atal and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", *The Journal of the Acoustical Society of America*, IEEE, vol. 50, no. 2, pp.637-655, 1971.

[13] M. J. Coker and S. F. Boll, "An improved isolation word recognition system based upon the linear prediction residual", *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '76*, vol. 1, pp. 206 – 209, 1976.

[14] L. R. Rabinar and M. R. Sambur, "Voiced-Unvoiced-Silence Detection Using the Itakura LPC Distance Measure", *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '77*, vol. 2, pp. 323-326.

[15] L. R. Rabinar and M. R. Sambur, "An algorithm for determining the endpoints of isolated utterances", The Bell System Technical Journal, pp. 297-315, 1975.

[16] S Furui, "Cepstral Analysis Technique for Automatic Speaker Verification", *IEEE*

Transactions on acoustics, speech and signal processing, vol. Assp- 29, no. 2, 1981

[17] M A Bush, G E Kopec and N Lauritzen, "Segmentation in Isolated Word Recognition Using Vector Quantization", Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84, vol. 9, 1984.

[18] L R Rabiner and B H Jaung, "An Introduction to Hidden Markov Models", *IEEE ASSP Magazine*, pp. 4-16, January 1986.

[19] H Harb, "Isolated word recognition using Neural Network", *Proceedings of The 7th IEEE International Conference on Electronics, Circuits and Systems, 2000. ICECS 2000*, Vol 1, 2000.

[20] S Axelrod, V Goel, R A Gopinath P A Olsen and K Vishweswariah, "Subspace Constrained Gaussian Mixture Models for Speech Recognition", *IEEE Transactions on speech and audio processing*, vol 13, no. 6, pp.1144-1160, November 2005.

[21] A R Mohamed, G E Dahl and G Hinton, "Acoustic Modeling Using Deep Belief Networks", IEEE Transaction on audio, speech and language processing, vol. 20, no. 1, January 2012.

[22] Ramana Rao G. V. and Srichand J, "Word boundary detection using pitch variations", *Proceedings of Fourth International Conference on Spoken Language, 1996. ICSLP 96*, pp. 813-816, May 1996.

[23] H A Patil and T K Basu, "Development of speech corpora for speaker recognition research and evaluation in Indian languages", *IJST 2008, Springer*, 2008.

[24] S Ranjan, "A Discrete Wavelet Transform Based Approach to Hindi Speech Recognition", *Proceedings of International Conference on Signal Acquisition and Processing, IEEE,* pp, 345-348, 2010.

[25] I Bhardwaj and N D Londhe, "Hidden Markov Model Based Isolated Hindi Word Recognition", *Proceedings of 2nd International Conference on Power, Control and Embedded Systems, IEEE,* 2012.

[26] T Pruthi, S Saksena and P K Das, "Swaranjali: Isolated Word Recognition for Hindi Language using VQ and HMM", *International Conference on Multimedia Processing and Systems (ICMPS),* IIT Madras.

[27] S Tripathy, N Baranwal and G C Nandi, "A MFCC based Hindi Speech Recognition Technique using HTK Toolkit", *Proceedings of the 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013*), pp. 539-544, 2013.

[28] H Sakoi and S Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", *IEEE Transactions on acoustics, speech and signal processing*, vol. Assp- 26, no. 1, February 1978.

[29] L R Rabiner, A E Rosenberg and S E Levinson, "Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition", *IEEE Transactions on acoustics, speech and signal processing*, vol. Assp- 26, no. 6, December 1978