Gujarati Speech Recognition – A Review

Rachana Bharat Parikh*1, Dr. Hiren Joshi *2
[1]Computer Science, Gujarat University, Ahmedabad, Gujarat, India,
rachanaparikh@gujaratuniversity.ac.in
[2]Computer Science, Gujarat University, Ahmedabad, Gujarat, India,
hirenjoshirajkot@gmail.com

Article Info
Volume 83
Page Number: 549 - 553
Publication Issue:
July - August 2020

Article History
***Article Received**: 06 June 2020*
***Revised**: 29 June 2020*
***Accepted**: 14 July 2020*
***Publication**: 25 July 2020*

Abstract

Automatic Speech Recognition is a techniques to determine human speech patterns, accordingly, relevant text output is obtained. Various interactive softwares are available in the market, however, due to vernacular languages the use of these softwares becomes limited as these software does not support all languages. In this paper we have reviewed Gujarati Speech Recognition System using different speech recognition methodologies which will cater the need of Gujarati users and will help them in technological advancement.

Keywords: Speech Recognition 1, Hidden Markov Model 2, Recurrent Neural Network 3, Deep Neural Network 4, End-to-End speech recognition 5

I. INTRODUCTION

Automatic Speech Recognition (ASR) is a technique to determine human voice. This is purely a process of converting analog signals into logical digital representation. In simple words, we can say that ASR processes and identifies human voice / words and provides us the logical output based on the spoken words / command. This process uses different variables and tools like viz. sound, statistics, speech variability, frequency digitization etc., to provide the desired output. Researchers, over a period of time have used the above tools and variables to dissect language and frequency of spoken words to provide various logical output. Basic features on which ASR depends includes the size of vocabulary, individual speaker, adaptation of speaker characteristics and the accent which is acoustic environment. Further, it also depends on the particular type of recognition of words are required to be done they may be, distinct word, sentence or natural dialect.

II. ASR METHODS USED FOR GUJARATI SPEECH RECOGNITION

A. Statistical

HMM is used to model the acoustic observations at the sub-word level, such as phonemes. Each model has to be typically modeled with 3 states. These three states includes separately model the beginning, middle and end of the phoneme. Each states evolves around its own transition as well as it may move to the next transition state. HMM will have a statistical distribution at each state which is also a mixture of diagonal co-variance Gaussian's. This will result in to likelihood for each of the vectors observed under the HMM model.

Acoustic vector is generated from speech audio signal. $Z_{1:T} = z_1, \ldots, z_T$ using feature extraction methods. From the feature vector, series of likely words are generated by decoder $w_{1:H} = w_1, \ldots, w_H$ which nearly closer to Z , wherein decoder tries to find

$$\hat{w} = \arg\max \{P(w|Z)\}.w \qquad (1)$$

P (w|Z) is difficult to compute directly so Bayes' Rule is used to change (1) into the similar problem of finding:

$$\hat{w} = \arg\max \{p(Z|w)P(w)\}.w \quad (2)$$

The probability p(Z |w) is ascertained by an acoustic model and the prior P(w) is ascertained by a language model.[4]

For example to recognize Gujarati numbers as isolated words individual word HMM are generated. *Word HMMs can be formed by concatenating its constituent phoneme HMMs.* Fig-1 show the word HMM of Gujarati number four (ચાર).

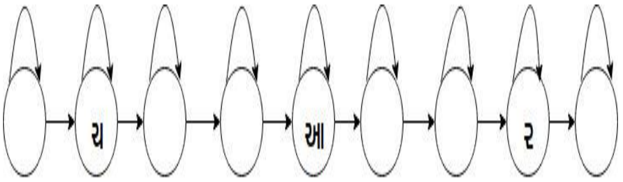

Fig 1: Example: ચાર triphone

Below Fig 2: shows how HMM is used to identify the isolated Gujarati number four (ચાર). From speech signal features are extracted using MFCC. From generated feature vector as observations the probability of likely word is calculated using formula (1).

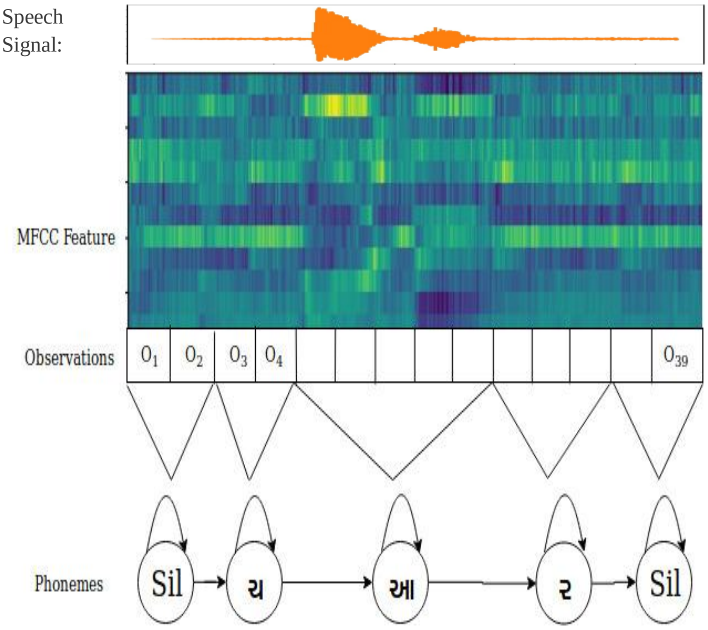

Fig 2: ચાર (Four) HMM generation process

Fig 3: shows how acoustic model, pronunciation dictionary and language model is used for decoding using search graph. Language model consist of probability of each words. Pronunciation dictionary consist of triphone with corresponding word. The HMM acoustic model phonemes are compared with HMM in dictionary and one having maximum score is considered as desired word.

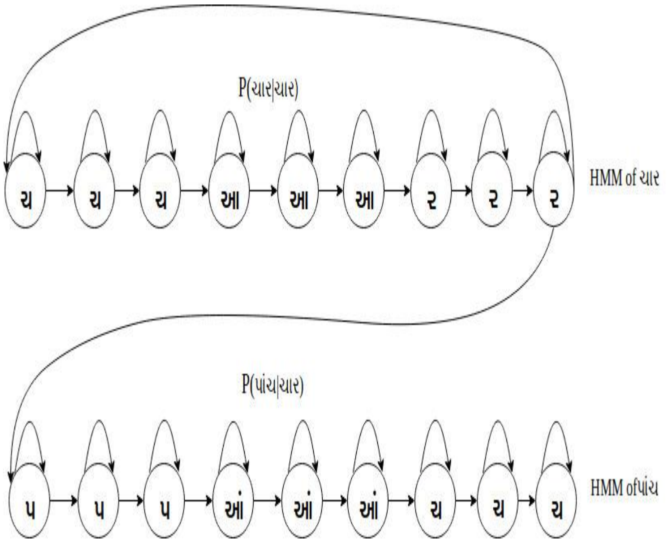

Fig 3: Search graph for ચાર and પાંચ (Four and Five)

B. Neural Networks

Artificial Neural Networks (ANN) is used in various phases of speech recognition including individual word recognition, phoneme identification, natural dialect recognition, etc, If ANN is compared to HMM modeling, ANN makes few explicit assumptions about feature statistical properties. Both HMM as well as ANN are combined together for getting better output as speech recognition requires both statistical comparison and better reasoning. Some of the neural networks used for Gujarati Speech Recognition are, Deep Neural Network (DNN) and Recurrent Neural Network(RNN) [2,3]. Fig. 4 show the example of gujarti speech recognition using RNN model.

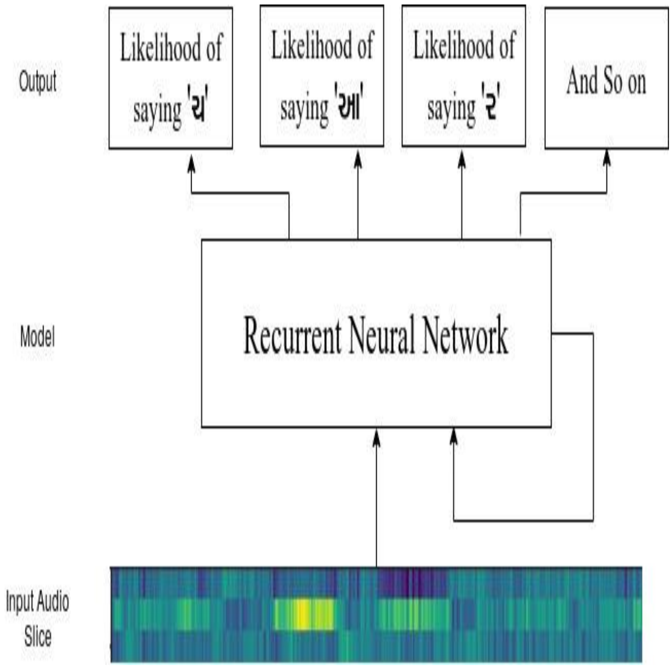

Fig 4: Example of Speech Recognition using RNN

C. End-to-End Model

End-to-End speech recognizer converts speech signal to characters without pronunciation dictionary, acoustic or language model. In Listen, Attend and Spell (LAS), the neural network model combine pronunciation dictionary, acoustic and language model together. LAS does not calculate statistical distribution of words based on acoustic features. LAS consists of listener and a speller. The listener is an encoder that accepts features as inputs and speller is a decoder that generate output character based on previous sequencing of characters [4].

III. Review of Gujarati ASR

A. Statistical

Jinal Tailor and Dipti Shah [5] have used Hidden Markov Model Toolkit HTK Tools for evaluating speech recognition accuracy. Their system perform on isolated words. They have used 10-12 words for testing. Their WER (Word Error Rate) was 5.85 % in ideal environment. They collected data for research from two groups: 4 male and 2 female varying in age range of 18 to 36.

Jinal Tailor and Dipti Shah [6] have used simple approach to recognize gujarati speech. They have used HMM technique. They generated speech corpus on their own which contained 650 regular Gujarati words which were recorded by 20 male and 20 female speakers from south Gujarat. They attain accuracy of 87.23% with WER 12.7%. They have used Viterbi algorithm for pattern matching to achieve better accuracy.

Mr. H. N. Patel and Dr. P. V. Virparia [1] have recognized 30 Gujarati words small vocabulary using English Speech Recognition Engine. They have used bootstrapping approach, which uses English speech recognition acoustic model for recognizing gujarati language phonemes. Following this approach required less training data. However, they require sound communication knowledge to create all possible rules for speech recognition.

B. Neural Networks

D. S. Pipalia Bhoomika Dave [13] designed a hybrid method (HMM/ANN) for isolated gujarati word recognition. System is designed in MATLAB. The data for research was recorded using Audacity software. The recording was done by 3 female and 2 male speaker. The data consists of 30 gujarati word including gujarati numbers and routine words. They achieved average accuracy of 70.57% and enhanced it to 79.14% by using Genetic algorithm.

H. B. Sailor, Mr. V. Siva Krishna [7] have done continuous Gujarati speech recognition. For front-end they have used Amplitude Modulation (AM) to extract features using the standard auditory filter banks. Time-Delay Neural Network (TDNN) and TDNN-Long Short-Term Memory (LSTM) models were used for acoustic modeling. They have used data set provided by SpeechOcean.com and Microsoft which is divided into a train and test sets. Recurrent Neural Network Language Models (RNNLM) is used.

S. Valaki and H. Jethva [9] implemented HMM/ANN for Gujarati ASR. They compared performance of HMM and HMM/ANN model on ten isolated Gujarati words. They used MATLAB for implementation. Some of the experiment showed improved performance in hybrid model. They calculated individual word recognition accuracy for many distinct Gujarati words.

P. Pravin and H. Jethva [10] they developed gujarati speech recognition using pattern recognition neural network. They identified 10 isolated gujarati words. The database was composed of 320 words as they recorded 10 words spoken by 4 speakers 8 times. A multilayered feed-forward Neural Network was used for speech recognition process. As gujarati is a tone language they selected a pre-emphasis factor which give importance to tone also along with vowels and consonants.

C. End-to-End Speech Recognition

H. K. Vydana, K. Gurugubelli, V. V. Raju, and A. K. Vuppala [8] They have used a joint acoustic model for multilingual ASR. Their study comprises of Subspace Gaussian mixture models (SGMM), and RNN with connectionst temporal classification (CTC) objective function. They concluded that RNN-CTC have performed better than SGMM system even on 120 hours of data (3 languages gujarati, tamil and telugu – 40 hrs each).

J. Billa [11] A multilingual training on three Indian languages: Telugu, Tamil and, Gujarati with a end-to-end LSTM based ASR system using CTC. They got reduced WER than challenge organizer's by 6.5% to 25.5%. Their multilingual training got additional 4.5% to 11.1% relative reduction in WER.

T. N. Sainath, R. J. Weiss, B. Li, P. Moreno, E. Weinstein, and K. Rao [12] A LAS model for sequence-to-sequence ASR model. They used joint grapheme sets of 9 languages and train model jointly on data from all languages. They achieved accuracy improvement by 21% relative compared to analogous sequence-to-sequence models trained on each language individually.

Work	Data Set	Method and Technique used	Research Work Done	Gujarati Speech RecognitionAccuracy Achieved
Isolated word	10 isolated word	HMM, HTK toolkit	Jinal Tailor and Dipti Shah [5]	95.1%
		HMM/ANN using MATLAB	S. Valaki and H. Jethva [9]	90.4%
	30 isolated word	Bootstraping	*Mr. H. N. Patel and Dr. P. V. Virparia* [1]	88.71%
		HMM/ANN	D. S. Pipalia Bhoomika Dave [13]	79.14%
Continuous speech	25 word sentence	HMM	Jinal Tailor and Dipti Shah [6]	87.23%
	Any number of words	TDNN with RNNLM	H. B. Sailor, Mr. V. Siva Krishna [7]	85.9%
Multilingual speech recognition using end-to-end model	Continuous speech of 3 indian language	RNN-CTC	H. K. Vydana, K. Gurugubelli, V. V. Raju, and A. K. Vuppala [8]	78.93%
	Continuous speech of 3 indian language	LSTM-CTC	J. Billa [11]	80.89%
	Continuous speech of 9 indian language	LAS model (Encode, Decoder, Attention model)	T. N. Sainath, R. J. Weiss, B. Li, P. Moreno, E. Weinstein, and K. Rao [12]	82.7%

TABLE – 1 PEER MODEL COMPARISON AND RESEARCH ADVANCEMENT

IV. CONCLUSION

This paper presents the research activities done in the area of Gujarati Speech Recognition using different platforms and experimenting the same on various models along with different sample sizes. The activity of speech recognition involves taking sound in the form of input and provide text, which exactly matches the sound. Speech recognition process deals with variability in individual's speech, range, pitch, accent, style of speaking etc. Researchers in their papers deliberated that Hybrid models like DNN/HMM increases the accuracy of recognition but this is done by using pronunciation dictionary. However, End-to-End model requires huge amount of data for speech recognition instead of pronunciation dictionary and provides more accurate results than other models as this model gets more combinations to deal before providing output.

V. REFERENCES

[1] H. N. Patel and Dr. P. V. Virparia, "A Small Vocabulary Speech Recognition for Gujarati," vol. 2, no. 1, 2011.

[2] A. L. Maas *et al.*, "Building DNN acoustic models for large vocabulary speech recognition," *Comput. Speech Lang.*, vol. 41, pp. 195–213, 2017.

[3] A. Graves, N. Jaitly, and A. Mohamed, "HYBRID SPEECH RECOGNITION WITH DEEP BIDIRECTIONAL LSTM Alex Graves , Navdeep Jaitly and Abdel-rahman Mohamed University of Toronto Department of Computer Science 6 King ' s College Rd . Toronto , M5S 3G4 , Canada," pp. 273–278, 2013.

[4] O. V. William Chan, Navdeep Jaitly, Quoc Le, "LISTEN , ATTEND AND SPELL : A NEURAL NETWORK FOR LARGE VOCABULARY CONVERSATIONAL SPEECH RECOGNITION Navdeep Jaitly , Quoc Le , Oriol Vinyals Google Brain," *1*, pp. 4960–4964, 2016.

[5] J. H. and D. B., "Speech Recognition System Architecture for Gujarati Language," *Int. J. Comput. Appl.*, vol. 138, no. 12, pp. 28–31, 2016.

[6] J. H. Tailor and D. B. Shah, "HMM-Based Lightweight Speech Recognition System for Gujarati Language," pp. 451–461, 2017.

[7] H. B. Sailor, M. V. Siva Krishna, D. Chhabra, A. T. Patil, M. R. Kamble, and H. A. Patil, "DA-IICT/IIITV system for low resource speech recognition challenge 2018," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2018-Septe, no. September, pp. 3187–3191, 2018.

[8] H. K. Vydana, K. Gurugubelli, V. V. V. Raju, and A. K. Vuppala, "An exploration towards joint acoustic modeling for Indian languages: IIIT-H submission for Low Resource Speech Recognition Challenge for Indian languages, INTERSPEECH 2018," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2018-Septe, no. September, pp. 3192–3196, 2018.

[9] S. Valaki and H. Jethva, "A hybrid HMM/ANN approach for automatic Gujarati speech recognition," *Proc. 2017 Int. Conf. Innov. Information, Embed. Commun. Syst. ICIIECS 2017*, vol. 2018-Janua, pp. 1–5, 2018.

[10]P. Pravin and H. Jethva, "Neural Network Based Gujarati Language Speech Recognition," vol. 2, no. May 2013, pp. 2623–2627, 2013.

[11]J. Billa, "ISI ASR system for the low resource speech recognition challenge for Indian languages," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2018-Septe, no. September, pp. 3207–3211, 2018.

[12]T. N. Sainath, R. J. Weiss, B. Li, P. Moreno, E. Weinstein, and K. Rao, "MULTILINGUAL SPEECH RECOGNITION WITH A SINGLE END-TO-END MODEL Shubham Toshniwal ∗ Toyota Technological Institute at Chicago," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 4904–4908, 2018.

[13]D. S. Pipalia Bhoomika Dave, "An Approach to Increase Word Recognition Accuracy in Gujarati Language," *Int. J. Innov. Res. Comput. Commun. Eng. (An ISO Certif. Organ.*, vol. 3297, no. 9, pp. 6442–6450, 2007.