ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

An Approach to Increase Word Recognition Accuracy in Gujarati Language

Bhoomika Dave, D. S. Pipalia

Department of Electronics and Communication, R.K. University, Rajkot, Gujarat, India

Assistant Professor, Department of Electronics and Communication, R.K. University, Rajkot, Gujarat, India

ABSTRACT: Speech recognition deals with identifying the spoken words and converting it into equivalent text form. Many application use speech recognition such as direct voice input in aircraft, data entry, speech-to-text processing, voice user interfaces such as voice dialing and many more. The paper presents a hybrid method for speech recognition(HMM/MLP) for isolated word recognition in Gujarati language. The feature extraction method used is MFCC. The implemented system is speaker independent and has been tested for 30 words in MATLAB environment.

KEYWORDS: Speech Recognition, MFCC, Hidden Markov model, Artificial neural network, HMM/MLP.

I. INTRODUCTION

The procedure of consequently perceiving talked expressions of speaker in light of data in Speech signal is called Speech Recognition. In Automatic Speech Recognition computers catches the words said by a human with an assistance of microphone. These words are then perceived via programmed speech recognizer, and at last, framework shows the perceived words on the screen. Speech processing can be performed at distinctive three levels. Signal level preparation considers the anatomy of human sound-related framework and process signal in type of little lumps called frames. In phoneme level handling, speech phonemes are gained and processed. Phoneme is the essential unit of Speech. Third level processing is known as word level handling. This model focuses on phonetic substance of speech. There are a couple of issues confronted in speech processing which can be recorded as Robustness, Adaptability, Language Modeling , Out of Vocabulary words, and Accent.

For the most part, there are three strategies generally utilized as a part of discourse acknowledgment: Dynamic Time Warping (DWT), Hidden Markov Model (HMM) and Artificial Neural Networks (ANNs). Dynamic Time Warping algorithm is utilized to recognize a segregated word sample equating it against various word formants to ascertain the particular case that best matches it. Dynamic Time Warping (DTW) is a productive system for discovering most favorable nonlinear arrangement between a format and the Speech sample. The primary issue of frameworks taking into account DTW is the little amount of learning words, high calculation rate and extensive memory necessity. [1]

A Hidden Markov Model is a statistical Markov Model in which the framework being demonstrated is thought to be a Markov process with unidentified (hidden) states. For speech recognition, the utilization of HMM is liable to the following restraints: Must be supported on a first order Markov chain, must have stationary states conversions; observations independence and likelihood limitations. Since speech recognition is fundamentally a pattern recognition issue, neural systems, which are great at pattern recognition, can be utilized for discourse recognition. Despite the fact that ANN has a decent discriminative force and adaptable it is not ready to give exceptionally precise calculation to successive information, like speech. [2]

This study conglomerates the advantages of the HMM and the ANN standards inside of a solitary framework to beat the constraints of any methodology working in isolation. The objective in this hybrid framework for ASR is to take advantage from the properties of both HMM and ANN enhancing its adaptability and recognition execution. A hybrid HMM/ANN recognizer that combines proficient discriminative learning capacities of NN and the prevalent time warping and decoding techniques connected with the HMM methodology was subsequently created. ANN was prepared to estimate HMM emission probabilities needed in HMM construct just in light of the acoustic data in a set

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

number of nearby speech frames. Those probabilities were then utilized by a Viterbi decoding procedure for recognition. [4] [5]

II. SYSTEM DESCRIPTION

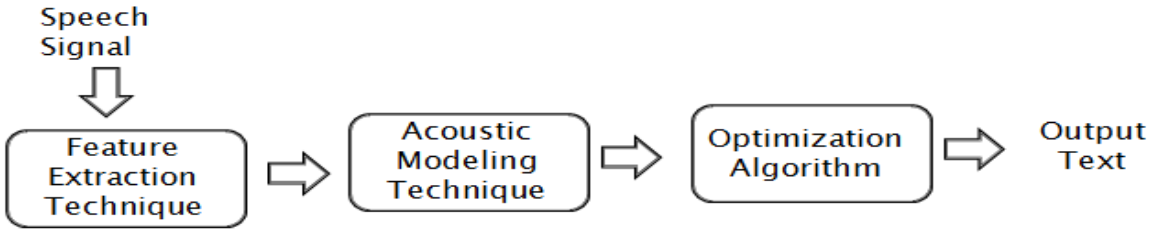

Figure 1: Block diagram for Speech Recognition

Speech recognition deals with the analysis of the linguistic content of a speech signal. A general speech recognition system may consist of five parts that is
1) Acquiring speech signal
2) Feature extraction
3) Modeling
4) Optimization if required.
The system is further described in detail with a short description of techniques used.

A) Speech signal:

These are the input signals provided to the system. The speech signals were recorded using a microphone in Audacity software. The acquired digital signal was then stored as .wav file for ease of manipulation and editing.

B) Feature Extraction :

Mel Frequency Cepstral Coefficients (MFCC) is a standout amongst the most regularly utilized element extraction system as a part of speech recognition taking into account recurrence area utilizing Mel scale which is in relative to human ear. MFCC is the representation of a cepstrum of windowed brief time sign got from FFT of that flag. MFCC utilizes a non straight recurrence scale which portrays the sound-related framework. MFCC is a component extraction procedure that concentrates parameters from speech that are utilized by human listening to furthermore de emphasizes every other parameter. The procedure of MFCC extraction is connected over every frame autonomously alluding to the figure demonstrated beneath the signal is initially isolated into time allotments comprising of a specific number of frames. The covering of the edges is utilized to smooth transition starting with one edge then onto the next. The filter bank is of overlapping triangular band pass filters which according to Mel-Scale provides linear equally spaced response for 1000 Hz and logarithmic beyond the 1000 Hz. The FFT signal obtained is then given to the filter bank. This signal obtained is provided to the logarithm block where Log-spectra-energy signal is obtained this energy signal is provided to the final block of DCT which gives desired set of MFFCs. One advantage of MFCC is that it is able to mimic Human Auditory System well. But it is very sensitive to noise and has no prediction algorithm. [14]

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

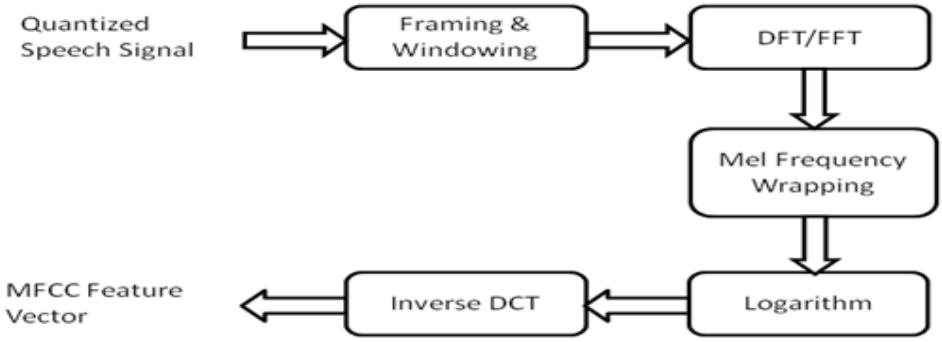

Figure 2: Steps for obtaining feature vectors by MFCC

C) Acoustic Modeling:

The modeling block comprises of acoustic modeling and dialect modeling. Acoustic model of framework models how a given word or "phone" is affirmed while the Language model predicts the probability of a given word showing up in the language (dictionary).The Modeling procedure utilized here is hybrid HMM/ANN .the essential point is to utilize the upsides of both the routines to invalidate the disservices that these frameworks independently confront. Registering word probabilities is done utilizing ANN.

A feed- forward neural network is utilized to figure the probabilities. The input quantities are spread across connections with other units. Every destination unit wholes its system input and processes an output value in range of 0 to 1 from this total. In the event that there are extra layers of units, the procedure is rehashed. In the end, the propagated initiation achieves the last layer of the system, which has no further associations. These units are the output of the net and, in a perfect word; their values speak of the likelihood that the information is from recorded word dictionary.

The search algorithm is completed utilizing HMM. Once the probabilities are processed, the framework utilizes a calculation called the Viterbi algorithm to locate the most elevated score way. The phoneme probabilities for each progressive casing are organized in a grid. At that point the framework finds the ways through the grids that gives the most astounding score and match the score with given target words. [13]

D) Optimization algorithm :

An optimization algorithm is a procedure which is executed iteratively by comparing various solutions till an optimum or a satisfactory solution is found. The Genetic algorithm is used here for providing optimization.

The optimization algorithm is a routine created to upgrade the outcomes got for the framework that is to show signs of improvement in the framework results. The advancement calculation utilized here is the Genetic algorithm which is quickly depicted beneath. The Genetic algorithm takes motivation from Darwin's Principle "Survival of the fittest".

The Genetic Algorithm is a transformative calculation. Here a representation arrangement is picked by the person to describe the plan of courses of action that can shape the database down the count. Specific number of arrangements is made to frame a beginning population. The following steps are then repeated time to time till a particular arrangement has been discovered which fulfills a pre-characterized end basis. Every single creature is judged utilizing a wellness capacity which is particular to the issue being understood. Based upon the wellness qualities acquired various creatures are decided to be parents. The new life forms (chromosomes) are created from those parents utilizing propagation administrators. The wellness estimations of that posterity are assessed. At last the survivors are chosen from the old population and the posterity to shape the new population of the present. The instruments that compute which and what number of parents to choose, what

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

number of changes to make, and which people will make into the cutting edge together address a decision method.[7]

The Block diagram for Basic Genetic Algorithm is as shown in figure given below and is described as follows.

- Starting is done with Genetic population (randomly selected population of n chromosomes) here considering a suitable solution for problem on hand.
- Then comes evaluating the fitness function of each chromosome in the population.
- Next step is creating new population by repeating the steps of Selection, crossover and Mutation.
    1. Selection: selecting any two chromosomes as parent chromosome from a population according to the values of fitness function.
    2. Crossover: taking crossover probability new offsprings can be obtained from crossover parents otherwise the offsprings are exact replica of parents.
    3. Mutation: taking mutational probability into consideration chromosomal alteration of new off spring at every set of point that are determined by given specific solutions.
    4. The loop ends when an end condition is satisfied else it goes back to step two.

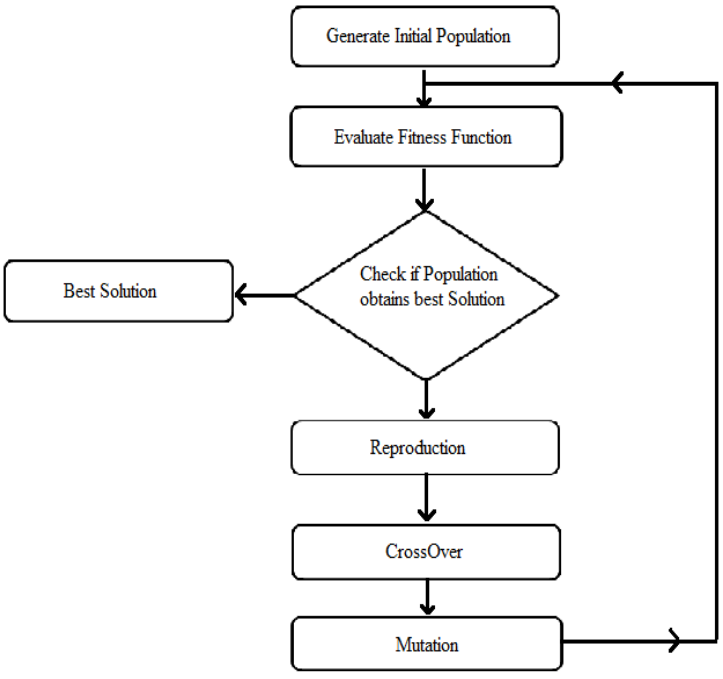

Figure 3 : Evolutionary process of Genetic algorithm

III. PROPOSED ALGORITHM

A. *Design Considerations:*

- Audacity used for recording speech signals.
- Mel Frequency Cepstral co-efficients (MFCC) for feature vectors (frames of word).
- ANN along with Genetic algorithm implementation.
- Hybrid HMM/MLP method used for pattern recognition.

B. *Description of the Proposed Algorithm:*

The database used here comprises of words recorded by five speakers where 3 female speakers and two male speakers were asked to record their speech. The recording was done through Audacity software. These recorded

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

speech signals were given to MFCC. MFCC was implemented with the help of rastamat and voice box toolkits. MFCC provided frames of a word these frames acted as an input to the Acoustic modeling block.

The number of frames for every word is utilized as inputs for the neural system. The vocabulary set is made out of thirty Gujarati words. Every word is spoken seven times by five speakers (two male and three female). Accordingly the database is made out of 1050 words. Frames got for every utterance of the speaker structure Mel-Frequency Cepstral Coefficients (MFCC). The database was partitioned into training, testing and validation sets. The training subset is utilized for processing the gradient and redesigning the system weights and biases. The validation subset is utilized to prevent model over fitting. The error on the validation set is observed amid the preparation process. Usually, the approval error decrease during the beginning period of the preparation as does the training set mistake. On the other hand, the validation set will normally start to rise when the system starts to over fit. On the off chance that the validation error keeps on expanding for a given number of epochs, the training will be halted. The testing subset is not utilized amid training , but rather it is utilized to compare models. [3]

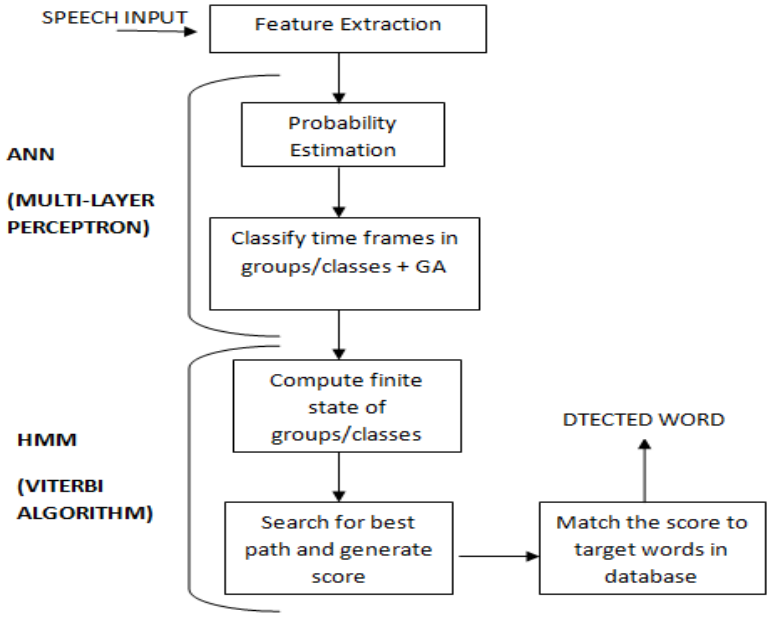

Figure 4 : Proposed system block diagram

A feed-forward neural networks that is Multi-layer Perceptron is used to compute word probabilities. The output values of certain units are set externally. These are the inputs to the network representing the frame of speech to be classified. These values are propagated across connections to other units. Each destination unit sums its network input and computes an output value in the range 0 to 1 from this sum. Eventually, the propagated activation reaches the final layer of the network, which has no further connections. These units are the output of the net and, ideally, their values represent the probability that the input is from words as in dictionary. The major difference between this approach compared to the standard HMM system is the posteriori probabilities are estimated using a neural network instead of a mixture of Gaussians. Using a neural network for estimation is better as it does not require assumptions on independence of the input data and can easily perform discriminative training. [13]

Once the phoneme probabilities are computed, the system use an algorithm called the Viterbi search to find the highest score path. The phoneme probabilities for each successive frame are arranged in a matrix. Then the system finds the paths through the matrixes that give the highest score and match the score with provided target words. Word scores that reflect relative merit are computed. The word score is compared with a threshold. If the score is better than the threshold, then accept the word. If worse, then reject the word. [13]

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

Further ahead the Genetic algorithm is implemented the working of which is explained in the earlier section.

English	Gujarati
One	એક
Two	બે
Three	ત્રણ
Four	ચાર
Five	પાઁચ
Six	છ
Seven	સાત
Eight	આઠ
Nine	નવ
Ten	દસ
Sky	આકાશ
Earth	ધરતી
Sea	દરિયો
Sun	સુરજ
Stars	તારા
Tulsi	તુલસી
Rose	ગુલાબ
Mogra	મોગરો
Python	અજગર
Crocodile	મગર
Chair	ખુરસી
Stairs	દાદરો
Terrace	અગાસી
Boon	વરદાન
River	નદી
Clouds	વાદળ
Water	પાણી
Gujarat	ગુજરાત
Wind	પવન
Vehicle	વાહન

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

IV. SIMULATION RESULTS

The experimental results for the four speakers are presented in the following figures. The database was divided into three groups for result generation. Where group 1 had all the digits spoken in Gujarati (1-10) group 2: had all the words recorded in Gujarati and listed in database table above and group 3 had the combined digits and words recorded in Gujarati making it a database of total 30 words each spoken seven times. The figures below shows performance chart for database of three sets that is 10 Digits, 20 Words, and 30 Digits plus words database.

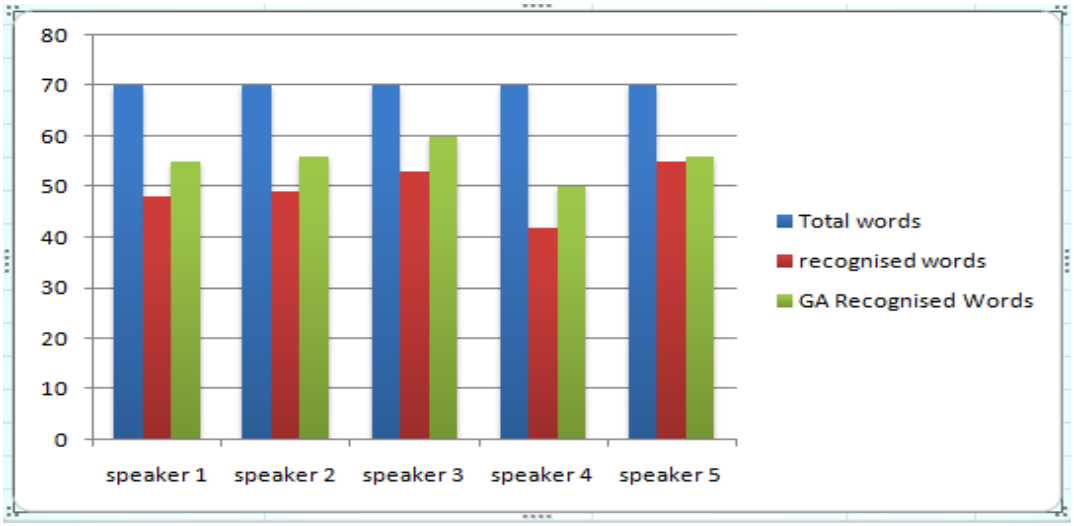

Figure 5 : Performance plot for database of 10 Digits

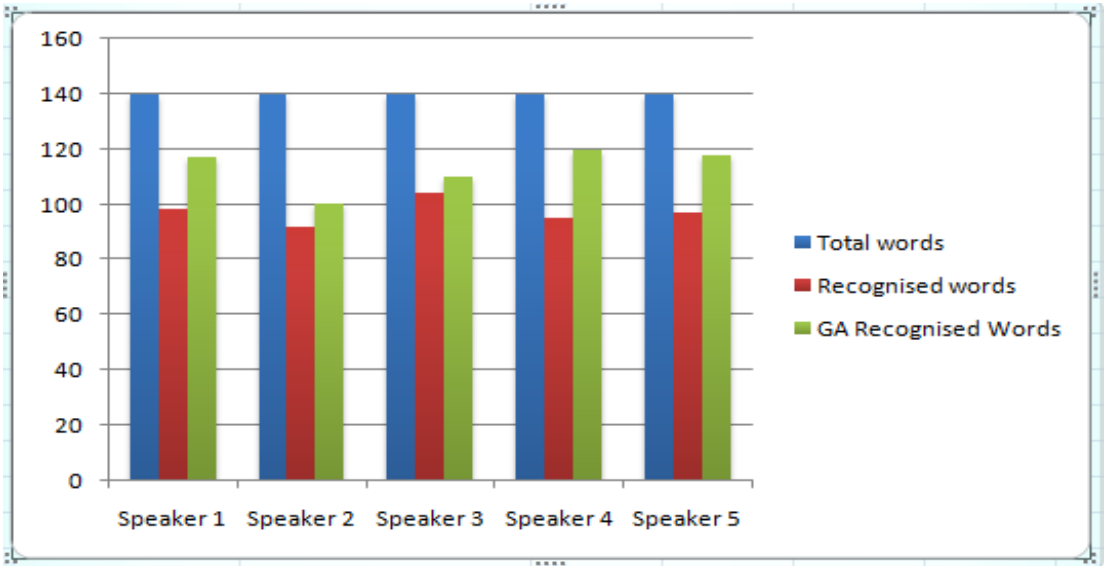

Figure 6 : Performance plot for database of 20 Words

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

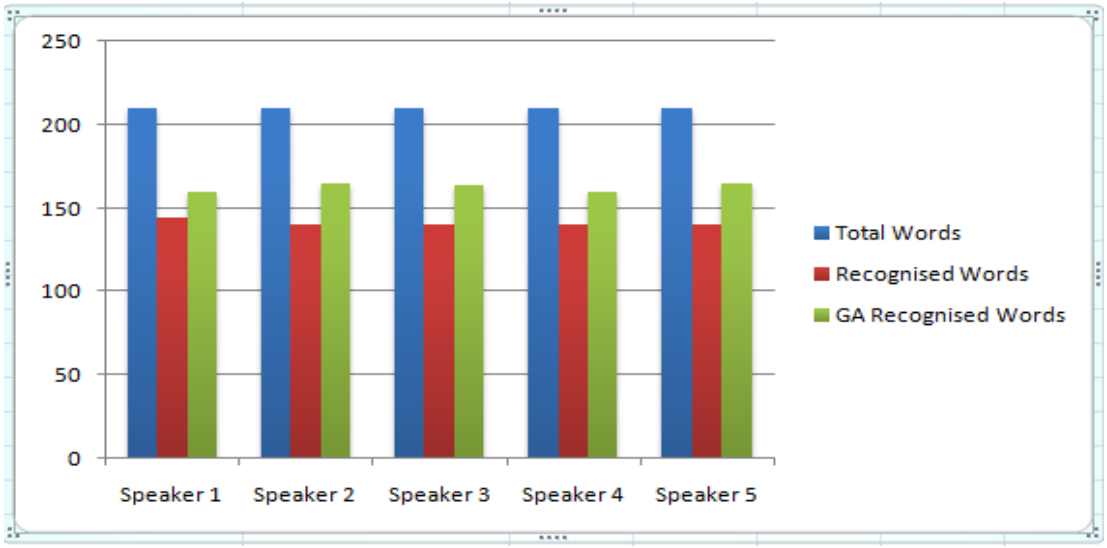

Figure 7: Performance plot for database of 30 words

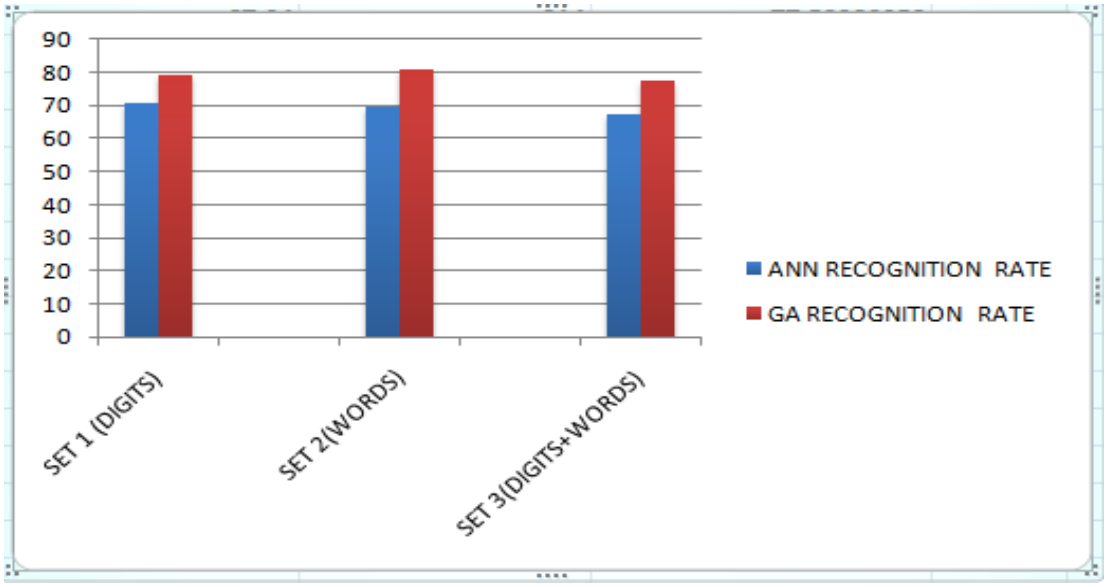

Figure 8 : Comparison of Recognition Rates obtained with ANN and ANN + GA

V. CONCLUSION AND FUTURE WORK

The paper describes hybrid method for speaker independent speech recognition in Gujarati language. The recognizer is implemented using MATLAB the words recorded were Gujarati digits (1-10) and certain words used in day to day life. The experimental results showed that the average accuracy for 10 speakers was 70.57% initially and increased to 79.14% after the implementation to Genetic algorithm. The results showed an optimal accuracy for recognition but it was also marked that the overall accuracy decreased as the number of words in the database increased. Gujarati being a tonal language the pitch differences prove of as much importance as the consonants and vowels so variants of the implemented system may be able to provide a better accuracy than obtained.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

REFERENCES

1) Lawrence R. Rabiner (February 1989). "A tutorial on Hidden Markov Models and selected applications in speech recognition". *Proceedings of the IEEE* **77** (2): 257–286.doi:10.1109/5.18626.
2) P.Patel, H. Jethva, 2013 "Neural Network Based Gujarati Language Speech Recognition" International Journal of Computer Science and Management Research Vol 2 Issue 5.
3) Mondher Frikha, 2012 "A Comparitive Survey of ANN and Hybrid HMM/ANN Architectures for Robust Speech Recognition" American Journal of Intelligent Systems.
4) Ajith Abraham "Artificial Neural Networks", Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn, John Wiley & Sons, Ltd. ISBN: 0-470-02143-8, 2005.
5) Herve Boulard, Nelson Morgan "Connectionist Speech Recognition A Hybrid Approach"KLUWER ACADEMIC PUBLISHERS, ISBN 0-7923-9396-1, 1994.
6) Hitesh Gupta, Deepinder Singh Wadhwa " Speech Feature Extraction and Recognition Using Genetic Algorithm" International Journal of Emerging Technology and Advanced Engineering Volume 4, Issue 1, January 2014 Pg: 363-369.
7) J. K. Patel, P. N. Patel, P. V. Virparia, 2013 " Voice Enabled Telephony Commands Using Gujarati Speech Recognition" International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 10.
8) Purnima Pandit, Shardav Bhatt 2014 "Automatic Speech Recognition of Gujarati digits using Dynamic Time Warping" International Journal of Engineering and Innovative Technology Volume 3, Issue 12.
9) Ms. Rupali S Chavan, Dr. Ganesh S. Sable 2013 "An Implementation of Text Dependent Speaker Independent Isolated Word Speech Recognition Using HMM" International Journal of engineering sciences & research technology.
10) Viresh Moonsar "Artificial Neural network based automatic speaker recognition using hybrid technique for Feature extraction".
11) Namrata Dave Article: Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition. *International Journal For Advance Research in Engineering And Technology* (ISSN 2320-6802) 07/2013; Volume 1(Issue VI).
12) H. F. Ong and A. M. Ahmad "Malay Language Speech Recogniser with Hybrid Hidden Markov Model and Artificial Neural Network (HMM/ANN)" International Journal of Information and Education Technology, Vol. 1, No. 2, June 2011 pg114-119.
13) Nidhi Desai, Prof.Kinnal Dhameliya, Prof.Vijayendra Desai, "Feature Extraction and Classification Techniques for Speech Recognition: A Review", International Journal of Emerging Technology and Advanced Engineering, (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 12, December 2013.