

Recognition of Spoken Gujarati Numeral and Its Conversion into Electronic Form

Bharat C. Patel

Smt. Tanuben & Dr. Manubhai Trivedi
. College of information science,
Surat, Gujarat, India,

Apurva A. Desai

Dept. of Computer Science, Veer Narmad South Gujarat
University, Surat, Gujarat, India

Abstract— Speech synthesis and speech recognition are the area of interest for computer scientists. More and more researchers are working to make computer understand naturally spoken language. For International language like English this technology has grown to a matured level. Here in this paper we present a model which recognize Gujarati numeral spoken by speaker and convert it into machine editable text of numeral. The proposed model makes use of Mel-Frequency Cepstral Coefficients (MFCC) as a feature set and K-Nearest Neighbor (K-NN) as classifier. The proposed model achieved average success rate of Gujarati spoken numeral is about 78.13%.

Keywords—speech recognition; MFCC; spoken Gujarati numeral; KNN

I. INTRODUCTION

Speech recognition is a process in which a computer can identify words or phrases spoken by different speakers in different languages and translate them into a machine readable-format. To do this task, vocabulary of words and phrases are required. Speech recognition software only identifies those words or phrases if they are spoken very clearly.

As per types of utterances a system can recognize, the speech recognition system is classified into two classes: Discrete Speech Recognition (DSR) system and Continuous Speech Recognition (CSR) system.

DSR system accepts pronunciation of a separate word, combination of words or phrases. Therefore, user has to make a pause between words as they were dictated. This system is also known as Isolated Speech Recognition (ISR) system.

CSR accepts pronunciation of continuous words. It uses special methods to determine utterance of word boundaries. It operates on speech in which words are connected together. i.e. not separated by pause. So, continuous speech is more difficult to handle than DSR.

The objective of this study is to build a speech recognition Interface/Tool for Gujarati language which helps people who are physically challenged to interact with computer. The proposed model allows user to speak Gujarati numeral via microphone and this spoken numeral is recognized by speech recognition tool and it is displayed into textual form.

A. Gujarati language

Gujarati is an Indo-Aryan language, descended from Sanskrit. Gujarati is the native language of the Indian state of

TABLE I. PRONUNCIATION OF EQUIVALENT ENGLISH AND GUJARATI NUMERALS.

English Digits	Pronunciation	Gujarati Numerals	Pronunciation
1	One	૧	Ek
2	Two	૨	Be
3	Three	૩	Tran
4	Four	૪	Chaar
5	Five	૫	Panch
6	Six	૬	Chha
7	Seven	૭	Saat
8	Eight	૮	Aath
9	Nine	૯	Nav
0	Zero	૦	Shoonya

Gujarat and its adjoining union territories of Daman, Diu and Dadra Nagar Haveli. Gujarati is one of the 22 official languages and 14 regional languages of India. It is officially recognized in the state of Gujarat, India. Gujarati has 12 vowels, 34 consonants and 10 digits. The pronunciation of ten English digits and their corresponding Gujarati numerals are given in Table I.

Gujarati is a syllabic alphabet in that all consonants have an inherent vowel. In fact, the very word 'consonant' means a letter that is pronounced only in the company of a 'vowel' sound. For instance, the Gujarati consonant 'સ' can be written, but it cannot be pronounced. If we want to pronounce this consonant we have to add any one of the vowels to it. Thus, if we add '૨' to 'સ' it becomes 'સ૨'. Thus, the pronunciation of the Gujarati numeral consists of both consonant and vowel. Therefore, it is difficult to recognize them easily. In this paper, we proposed a model that recognize all Gujarati numerals, i.e., ૦-૯.

B. Challenges in identification of spoken Gujarati numeral

There is no or little work done in Gujarati language on identification of spoken Gujarati numeral. This is our first effort to develop an interface that recognizes spoken Gujarati numeral. During the study of our work, we may find some of

the problems which create an ambiguity to recognize spoken Gujarati numerals. Let us discuss the different circumstances which create confusion to recognize spoken Gujarati numerals:

- Dissimilar pronunciation of same numeral by same speaker in various situations.
- Dissimilar pronunciation of same numeral by different speakers.
- Pronunciation of Gujarati numeral is not clear or may include background noise.
- When each Gujarati consonant is pronounced, it is succeeded by a vowel.
- The pronunciation of a speaker from different districts of Gujarat state also differs.

Because of these problems, the recognition of spoken Gujarati numeral is more complicated. So, it requires some additional action to be applied on it rather than other languages.

This paper has basically six sections. The introductory section is followed by related work. The third section shows our proposed model for recognition of spoken Gujarati numeral and the fourth section enumerates the methodology proposed. In the next section of the paper the results are shown, which are derived by our experiments and finally, the conclusion is given.

II. RELATED WORK

In this section, overview of some of the research works related to speech recognition for national and international languages is given.

In 2010, Patel and Rao [1] presented a paper on the recognition of speech signal using frequency spectral information with Mel frequency for the improvement of speech feature representation in HMM based recognition approach. Nehe and Holambe [2] proposed a new efficient feature extraction method using Dynamic Time Warping (DWT) and Linear Predictive Coding (LPC) for isolated Marathi digits recognition. Their experimental result shows that the proposed Wavelet sub-band Cepstral Mean Normalized (WSCMN) features yield better performance over Mel-Frequency Cepstral Coefficients (MFCC) and Cepstral Mean Normalization (CMN) and also give 100% recognition performance on clean data. The feature dimension for WSCMN is almost half of the MFCC. This reduces the memory requirement and the computational time. Pour and Farokhi [3] presented an advanced method which is able to classify speech signals with the high accuracy of 98% at the minimum time. Al-Alaoui M. A. et. al. [4] compare two different methods for automatic Arabic speech recognition for isolated words and sentences. The speech recognition system is implemented as a part of the Teaching and Learning using Information Technology (TLIT) project which would implement a set of reading lessons to assist adult illiterates in developing better reading capabilities. The first stage involved the identification of the different alternatives for the different components of a speech recognition system, such as using linear predictive coding, using HMMs, Neural Networks (NN) or KNN Classifier for the pattern recognition block. They

trained NN classifier using the Al-Alaoui algorithm overcomes the HMM in the prediction of both words and sentences. They also examined the KNN classifier which gave better results than the NN in the prediction of sentences. Al-Haddad S.A.R. et. al. [5] presented a pattern recognition fusion method for isolated Malay digit recognition using DTW and HMM. This paper has shown that the fusion technique can be used to fuse the pattern recognition outputs of DTW and HMM. Furthermore, it also introduced refinement normalization by using weight mean vector to get better performance with accuracy of 94% on pattern recognition fusion HMM and DTW. Rathinavelu A. et. al. [6] developed Speech Recognition Model for Tamil Stops. The system was implemented using Feedforward neural networks (FFNet) with backpropagation algorithm. This model consists of two modules, one is for neural network training and another one is for visual feedback and an average accuracy level of 81% has been achieved in the experiments conducted using the trained neural network. El-obaid M. et. al. [7] presented their work on the recognition of isolated Arabic speech phonemes using artificial neural networks and achieved a recognition rate within 96.3% for most of the 34 phonemes. Yamamoto K. et. al. [8] proposed a novel endpoint detection method which combines energy-based and likelihood ratio-based Voice Activity Detection (VAD) criteria, where the likelihood ratio is calculated with speech/non-speech Gaussian Mixture Models (GMMs). Moreover, the proposed method introduces the Discriminative Feature Extraction (DFE) technique in order to improve the speech/non-speech classification. Pinto and Sitaram [9] proposed two Confidence Measures (CMs) in speech recognition: one based on acoustic likelihood and the other based on phone duration and have a detection rate of 83.8% and 92.4% respectively. Bazzi and Katabi [10] presented a paper on recognition of isolated spoken digits using Support Vector Machines (SVM) classifier. They achieved 94.9% accuracy using SVM classifier. Patel and Desai [11] presented a paper on recognition of isolated spoken Gujarati numeral model which uses MFCC feature extraction method and DTW classification and achieved average accuracy rate of 71.17% for Gujarati numerals.

III. PROPOSED MODEL FOR RECOGNITION OF SPOKEN GUJARATI NUMERAL

Our proposed speech recognition model work only for Gujarati numerals. This model is an isolated word, speaker independent speech recognition system which uses template based pattern recognition approach. The Fig. 1 shows the block diagram of proposed model which recognizes isolated Gujarati numerals spoken by different speakers. The model consists of mainly three components: digitization, feature extraction, and pattern classification.

Practically, the function of digitization stage is to acquire analog signal of spoken numeral produced by person via microphone and convert them into digital signal. This digitized signal is conveyed to the next stage of proposed model named feature extraction, heart of proposed model. The model uses a MFCC (Mel-Frequency Cepstrum Co-efficient) as feature extraction method which accept digital signal and generates a feature vector of spoken Gujarati numeral. MFCC includes intermediate steps such as framing, windowing, Fast Fourier Transform (FFT), Mel Frequency wrapping and

finally computing the DCT (Discrete Cosine Transform) to produce feature vector of spoken numeral. Framing is the segmentation of the speech wave in which the speech signal is assumed to be stationary with constant statistical properties. Hamming window is used to decrease the signal to zero at the beginning and end of each frame. Then FFT is used to convert each frame of N samples from the time domain into the frequency domain.

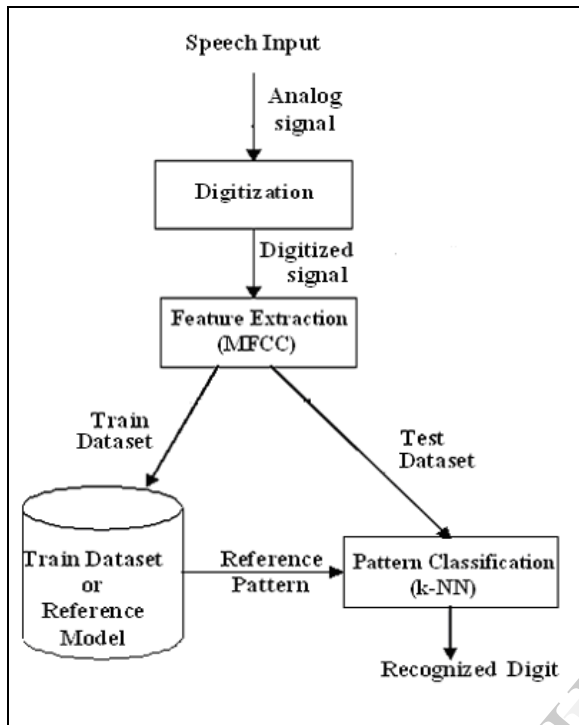


Fig. 1. Block diagram of proposed model.

The Mel-frequency Wrapping is used to obtain a mel-scale spectrum of the signal from the frequency domain. In the final step, the log mel spectrum is converted back to time domain and the result is called the mel frequency cepstrum coefficients, i.e. MFCC.

IV. METHODOLOGY

In this work, we have collected speech samples of all Gujarati numerals spoken by different speakers. Speech samples are mostly concerned with recording speech of each Gujarati numerals, ૦-૯, pronounced by different speakers. We consider four main factors while collecting speech samples, which affect the training set vectors that are used to train the data set. The first factor is the profile of the speakers which consists of range of age and gender of speakers. For proposed model, we have taken speech samples of 600 speakers, among them 50% are male speakers and 50% are female speakers, belonging to heterogeneous as well as homogeneous age groups. The second factor is the speaking conditions, i.e. the environment in which the speech samples were collected from. Here, we collected speech samples of Gujarati numeral not in a quiet or noise proof environment, it means that all the speech samples were interrupted by noise. The basis behind collecting the speech samples from noisy environment is to represent a real world speech samples

collection, because most speech recognition systems are intended to be used in different environment. Therefore, collecting speech samples from noisy environment was purposely done. The third factor is the transducers and transmission systems. In this work, speech samples were recorded and collected using a normal microphone. The fourth factor is the speech units. The system's main speech units are Gujarati spoken numerals, that means zero (૦) to nine (૯).

We have developed MATLAB GUI interface which records Gujarati numeral utterance produced by speaker through a microphone. This utterance is passed on to the Feature Extraction module. The feature extraction module extracts the unique feature of spoken data using feature extraction method known as MFCC. The mel value for given frequency f is calculated using Eq. (1) as given below:

$$F_{\text{mel}}(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (1)$$

In feature extraction stage, we computed matrix of mel filter coefficient, compute mel spectrum from time signal and finally constructed mat file which contains features of spoken Gujarati numeral. A length of feature vector of each spoken Gujarati numeral is 3234. These features are stored in database, known as train dataset or reference model. For pattern classification, according to Desai [12] different types of classifier like template matching, artificial neural network, K-nearest Neighbor (K-NN) are available and experimented by various researchers. In the classification phase K-NN classifier is used to classify test pattern of spoken numeral. Here, reference patterns stored in reference model are compared with test pattern. K-NN classifier uses Euclidean distance measure to find the nearest match between train and test pattern. If spoken data (i.e. test pattern) is matched with reference pattern, then the proposed model translate them into textual numeral and display on the speech conversion window.

V. EXPERIMENTAL RESULTS

The speech utterances were not recorded in a quiet or noise proof room. The speech duration to record isolated Gujarati numeral is 1.5 seconds and frequency sampling rate was 8 kHz. To evaluate the performance of the proposed model, the speech material used in the experiment was a speech sample of spoken Gujarati numeral database produced by 600 speakers of heterogeneous age groups. Each speaker pronounced 10 Gujarati numerals, i.e. ૦-૯. So that, the total number of speech samples is 6000.

For experiment purpose, we created two types of datasets namely train dataset and test dataset. Further, as per the age of speakers, they are categorized into two types: (i) heterogeneous age group of speakers and (ii) homogeneous age group of speakers.

The accuracy rate of individual spoken Gujarati Numeral is calculated using Eq. (2) as follow:

$$\text{Accuracy rate}(\%) = \frac{S}{T} \times 100 \quad (2)$$

Where S = Number of successful detection of test digit

T = Number of digits in the train dataset.

Moreover, average accuracy rate of all Gujarati numerals is calculated by taking the sum of accuracy of each numerals divided by 10.

A. Heterogeneous age group of speakers

In this work, experiment carried out on speech samples of heterogeneous age of speakers having age range between 5 and 40 years.

TABLE II. ACCURACY RATE OF GUJARATI NUMERALS FOR TRAIN AND TEST DATA SET OF SIZE 250

Test Numerals	Train Numerals										Acc.(%)	Missed
	૦	૧	૨	૩	૪	૫	૬	૭	૮	૯		
૦	210	5	13	8	3	0	0	0	2	9	84.00	40
૧	5	200	32	3	7	0	1	0	0	2	80.00	50
૨	4	17	211	3	1	3	1	1	1	8	84.40	39
૩	1	1	1	181	14	3	8	9	11	21	72.40	69
૪	0	3	3	32	120	4	10	48	30	0	48.00	130
૫	0	4	0	21	10	149	2	11	50	3	59.60	101
૬	1	1	2	37	7	2	183	10	1	6	73.20	67
૭	0	1	0	32	51	8	21	106	28	3	42.40	144
૮	0	0	0	47	17	28	3	16	137	2	54.80	113
૯	1	1	6	18	5	0	10	0	2	207	82.80	43

Initially, we have taken speech samples of 500 speakers among them 250 speech samples are used for train dataset and 250 speech samples are used for test dataset. The proposed model is applied on these dataset.

Table II shows the accuracy rate of each test Gujarati numerals against train Gujarati numerals. Let us examine the results obtained for numeral zero (૦). The finding in table II indicates that test numeral zero (૦) successfully matched with train numeral zero (૦) 210-times. In other words, out of 250 test numerals of zero, 40 numerals are not matched with train numerals. Because, it matches 5-times with numeral one (૧), 13-times with numeral two (૨), 8-times with numeral three (૩), 3-times with numeral four (૪), 2-times with numeral eight (૮) and 9-times with numeral nine (૯). Therefore, accuracy rate of test numeral zero (૦) is calculated using Eq. (2) as follow:

$$\begin{aligned} \text{Accuracy rate of test numeral zero (\%)} &= 210 * 100/40 \\ &= 84.00\% \end{aligned}$$

Likewise, we can calculate accuracy rate for rest of the numerals. Let us examine the accuracy rate of each numeral. Numerals zero (૦), one (૧), two (૨), three (૩), six (૬) and nine (૯) achieved success rate more than 70%, numerals five (૫) and eight (૮) achieved more than 55% and numerals four (૪) and seven (૭) achieved less than 50%. The over all average accuracy rate of all numerals is 68.16 %.

Moreover, we have taken speech samples of 600 speakers and created two datasets train and test of equal number of speech samples, i.e. 300 speech samples per dataset. The outcome of table III denotes the accuracy rate of each test Gujarati numerals. The accuracy rate of numerals zero (૦), one (૧), two (૨), six (૬) and nine (૯) is more than 80%, numerals three (૩), five (૫) and eight (૮) is more than 70% and numerals four (૪) and seven (૭) is less than 70%. Here, we achieved over all average accuracy rate of all Gujarati numerals is 78.13% which is greater than average accuracy rate obtained for all numerals in table II.

It should be obvious from the results obtained in table II and table III that the accuracy rate of individual numerals and average accuracy of all numerals is increased when we increase speech samples in train and test datasets.

Also, we have applied proposed model on unequal size of both the datasets i.e. train and test dataset. In this work, we have taken speech samples of 600 speakers and created two datasets of unequal size i.e., out of 600 speech samples, 350 speech samples are used for train dataset and 250 speech samples are used for test dataset. Table IV enumerates the accuracy rate of individual test Gujarati numerals. Here, all Gujarati numerals achieved accuracy rate more than 70% which is better result than equal size of datasets. Moreover, some of the numerals achieved success rate nearer or more than 90%. The average accuracy rate of all Gujarati numerals, i.e. ૦-૯, is 80.84 %.

TABLE III. ACCURACY RATE OF GUJARATI NUMERALS FOR TRAIN AND TEST DATA SET OF SIZE 300

Test Numerals	Train Numerals										Acc.(%)	Missed
	૦	૧	૨	૩	૪	૫	૬	૭	૮	૯		
૦	262	2	15	3	1	0	4	1	0	12	87.33	38
૧	6	242	44	2	2	0	1	0	0	3	80.67	58
૨	4	17	263	1	3	0	1	1	2	8	87.67	37
૩	2	0	1	215	9	9	20	12	13	19	71.67	85
૪	0	0	1	19	202	6	14	35	23	0	67.33	98
૫	1	0	1	13	9	219	2	17	34	4	73.00	81
૬	2	1	0	11	16	1	254	8	1	6	84.67	46
૭	0	0	0	14	39	7	13	205	19	3	68.33	95
૮	0	0	1	22	16	26	3	17	214	1	71.33	86
૯	1	0	7	11	1	0	10	1	1	268	89.33	32

TABLE IV. ACCURACY RATE OF GUJARATI NUMERALS FOR TRAIN AND TEST DATA SET OF SIZE 350 AND 250 RESPECTIVELY

Test Numerals	Train Numerals										Acc.(%)	Missed
	૦	૧	૨	૩	૪	૫	૬	૭	૮	૯		
૦	221	2	11	1	3	0	4	1	0	7	88.40	29
૧	2	206	35	2	2	0	1	0	0	2	82.40	44
૨	3	12	225	0	1	0	1	1	1	6	90.00	25
૩	0	0	1	186	7	8	15	6	11	16	74.40	64
૪	0	0	0	12	177	4	8	32	17	0	70.80	73
૫	0	0	2	10	4	192	2	13	26	1	76.80	58
૬	1	1	0	9	11	1	216	6	1	4	86.40	34
૭	0	0	0	11	33	5	10	175	14	2	70.00	75
૮	0	0	1	15	11	19	1	10	192	1	76.80	58
૯	0	0	4	6	0	0	7	1	0	232	92.80	18

The total number of speech samples in table III and table IV are same but the size of train and test dataset in table III are same, i.e. 300 speech samples, whereas in table IV it is different, i.e., speech samples in train and test dataset is 350 and 250 respectively. Their finding shows that the accuracy rate of individual test numerals and average accuracy is increased if speech samples are increased in train dataset.

We have also experimented on different size of dataset such as 50, 100, 150, 200, 250 and 300 for both train as well test dataset and found average accuracy rate of each dataset is 47%, 53.70%, 62.47%, 66.35%, 72.40% and 78.13% respectively. These results show that when we increase size of dataset, the average accuracy rate also increases. In other words, if there is more number of speech samples then the performance of the system is increased.

B. Homogeneous age group of speakers

In this work, we experimented on speech samples of homogeneous age group of speakers. As per the age of speaker, they are further sub-divided into two groups: (a) speech samples of speakers of age between 5 and 15 years and (b) speech samples of speakers of age between 16 and 40 years. We created train and test datasets for both the groups and then proposed model is applied on these datasets.

1) Speech samples of speakers of age between 5 and 15 years

In this group, we have included speech samples of 210 speakers, each speakers belonging to age range between 5 and 15 years, among them 105 speech samples are used as train

dataset and 105 speech samples are used as test dataset. Table-V shows the accuracy rate of individual Gujarati numerals and the over all average accuracy of all numerals is 64.48 %

The finding show that accuracy rate of all Gujarati numerals for homogeneous age group of speakers is better than the heterogeneous age group of speakers of same size dataset.

2) Speech samples of speakers of age between 16 and 40 years

It is another group of homogeneous age of speakers. This group includes speech samples of those speakers whose age is between 16 and 40 years. We have taken speech samples of 310 speakers, out of them half of the speech samples are used for train dataset and half of the speech samples are used for test dataset. Table-VI shows the accuracy rate of each Gujarati numerals. Total average accuracy of all numerals is 60.45%

It is obvious from result obtained in table V and table VI that accuracy rate of Gujarati numerals for the speakers of age group between 5 and 15 years is higher than the speakers of age group between 16 and 40 years. In other words, we can say that the younger group of speakers have higher accuracy rate than the older group of speakers.

There is no literature found for spoken Gujarati numerals. However, we found some literature for national and international languages to compare the results of this work. Singh and Kumar [13] presented their work on recognition of isolated spoken word of English language. Authors achieved

the accuracy percentage 95% for the trained set and 81.23% for the test set.

TABLE V. ACCURACY RATE OF NUMERALS FOR TRAIN AND TEST DATASET OF SIZE 105

Test Numerals	Train Numerals										Acc.(%)	Missed
	૦	૧	૨	૩	૪	૫	૬	૭	૮	૯		
૦	82	2	10	2	0	0	2	1	0	6	78.10	23
૧	2	77	17	1	3	1	0	1	0	3	73.33	28
૨	2	11	84	3	2	1	0	0	0	2	80.00	21
૩	2	0	1	63	7	5	8	4	4	11	60.00	42
૪	0	0	0	11	48	4	9	19	14	0	45.71	57
૫	1	0	0	9	5	55	0	12	21	2	52.38	50
૬	0	0	1	8	6	0	78	8	2	2	74.29	27
૭	0	0	2	8	21	8	6	47	11	2	44.76	58
૮	0	0	1	11	10	11	1	9	60	2	57.14	45
૯	1	0	3	8	3	0	5	1	1	83	79.05	22

TABLE VI. ACCURACY RATE FOR TRAIN AND TEST DATA SET OF SIZE 155

Test Numerals	Train Numerals										Acc.(%)	Missed
	૦	૧	૨	૩	૪	૫	૬	૭	૮	૯		
૦	123	4	9	4	4	0	1	4	0	6	79.35	32
૧	8	103	28	4	7	0	0	2	3	0	66.45	52
૨	7	15	112	3	5	1	2	1	1	8	72.26	43
૩	1	1	2	107	9	0	6	14	5	10	69.03	48
૪	0	1	1	28	67	4	8	33	12	1	43.23	88
૫	0	2	0	20	6	83	3	18	22	1	53.55	72
૬	0	2	0	29	9	1	101	7	1	5	65.16	54
૭	0	1	2	28	29	4	13	65	13	0	41.94	90
૮	0	0	1	36	12	15	0	20	70	1	45.16	85
૯	2	1	5	17	4	0	15	2	3	106	68.39	49

Pour and Farokhi [3] presented a new method that developed an automatic Persian speech recognition system. They achieved the system accuracy rate to 98%. Alotaibi Y. A. et. al. [14] designed a system to recognize an isolated whole-word speech for Arabic digits zero to nine. This recognition system achieved 93.72% overall correct rate for Arabic digit. Al-Haddad S.A.R. et. al. [5] presented a pattern recognition fusion method for isolated Malay digit recognition using Dynamic Time Warping (DTW) and Hidden Markov Model (HMM). They obtained about 80.5% of accuracy for DTW and 90.7% for HMM and 94% for pattern recognition fusion DTW-HMM method. Rathinavelu A. et. al. [6] developed Speech Recognition Model for Tamil Stops. They obtained an average accuracy rate 81% in the experiments conducted using the trained neural network. Deemagarn and Kawtrakul [15] presented speech recognition system of speaker-independent Thai connected digit. The average recognition rate is 75.25 % for known length strings and 70.33% for unknown length strings.

VI. CONCLUSION AND FUTURE WORK

In this work, the proposed model accepts Gujarati numeral spoken by speaker via microphone and then that spoken numeral is converted into editable text by MATLAB. The proposed recognition system was experimented on speech samples of different size of datasets. Speech samples of heterogeneous group of speakers for train and test dataset of

size 300, the overall system performance of proposed model is 78.13%. The experimental result shows that the performance of the proposed system is improved for all Gujarati numerals when more train speech samples are used, i.e., greater the train speech samples, greater the chances of accuracy rate for Gujarati numerals. For homogeneous group, the accuracy rate is achieved 64.48% for the speaker of age group between 5 and 15 years with dataset size 105 and 60.45% for the speaker of age group between 16 and 40 years with dataset size 155. The experimental result of the proposed system achieved good result for homogeneous group of speakers having age between 5 and 15 years than the other two groups such as heterogeneous group and homogeneous age group of speakers having age between 16 and 40 years. So, we can declare that, if there are more speech patterns of younger speakers then the efficiency of the system is increased. The best correct rates were encountered in the case of most of the Gujarati numerals, but the worst correct rates were encountered in the case of numerals four (૪) and seven (૭) for both heterogeneous as well as homogeneous group.

The proposed algorithm works only for isolated Gujarati numeral. In future, we can modify this algorithm for continuous spoken Gujarati numeral as well as for isolated and continuous spoken Gujarati words.

REFERENCES

- [1] Patel Ibrahim, Rao Y. S., "Speech Recognition using HMM With MFCC- An Analysis using Frequency Spectral Decomposition Technique", *Signal & Image Processing : An International Journal(SIPIJ)* Vol.1, No.2, 2010, pp. 101-110.
- [2] Nehe N S., Holambe R. S., "New Feature Extraction Techniques for Marathi Digit Recognition", *International Journal of Recent Trends in Engineering*, Vol 2, No. 2, 2009, pp. 22-24.
- [3] Pour M. M., Farokhi F., "An Advanced Method for Speech Recognition", *World Academy of Science, Engineering and Technology*, 49, 2009, *Engineering Education Magazine*, Vol. 3, No. 3, 2008, pp. 77-86.
- [4] Al-Alaoui M. A., Al-Kanj L., Azar, J., Yaacoub E., "Speech Recognition using Artificial Neural Networks and Hidden Markov Models", *IEEE Multidisciplinary* pp. 995-1000.
- [5] Al-Haddad S. A.R., Samad S. A., Hussain A., Ishak, K. A., "Isolated Malay Digit Recognition Using Pattern Recognition Fusion of Dynamic Time Warping and Hidden Markov Models", *American Journal of Applied Sciences*, Vol. 5, No. 6, 2008, pp. 714-720.
- [6] Rathinavelu A., Anupriya G., Muthananthamurugavel A.S., "Speech Recognition Model for Tamil Stops", *Proceedings of the World Congress on Engineering*, Vol. 1, 2007, pp. 128-131
- [7] El-Obaid M., Al- Nassiri A., Maaly I. A., "Arabic Phoneme Recognition Using Neural Networks", *Proceedings of the 5th WSEAS International Conference on Signal Processing*, 2006, pp. 99-104.
- [8] Yamamoto K., Jabloun F., Reinhard K., Kawamura A., "Robust Endpoint Detection for Speech Recognition Based on Discriminative Feature Extraction", Vol. 1, 2006, pp. 805-808.
- [9] Pinto J., Sitaram R.N.V., "Confidence Measures in Speech Recognition based on Probability Distribution of Likelihoods", *Interspeech*, Lisbon, Portugal, 2005, pp. 3001-3004.
- [10] Bazzi I., Katabi D., "Using Support Vector Machines for Spoken Digit Recognition", In *Proceedings of INTERSPEECH*, 2000, pp. 433-436.
- [11] Patel B. C., Desai A. A., "Recognition of spoken Gujarati numerals using Dynamic Time Warping", *VNSGU Journal of Science and Technology*, Vol. 3, No. 2, 2012, pp. 81-88.
- [12] Desai A. A., "Handwritten Gujarati Numeral Optical Character Recognition using Hybrid Feature Extraction Technique", *International Conference IP, Comp. Vision and Pattern Recognition*, 2010, pp. 733 – 739.
- [13] Singh G. S., Kumar D., "Isolated word Recognition System for English Language", *International Journal of Information Technology and Knowledge Management*, Volume 2, No. 2, 2010, pp. 447-450.
- [14] Alotaibi Y. A., Alghamdi M., Alotaiby F., "Speech Recognition System of Arabic Digits based on A Telephony Arabic Corpus", *The International Conference on Image Processing, Computer Vision, and Pattern Recognition*, 2008, pp. 14-17.
- [15] Deemagarn A., Kawtrakul A., "Thai Connected Digit Speech Recognition Using Hidden Markov Models", *SPECOM'2004: 9th Conference Speech and Computer*, 2004, 20-22.

IJERT