Volume 5, Issue 11, November 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Spoken Digit Recognition Based on Support Vector Machine for Kannada Language

S B Harisha
Dept of TCE, JNNCE
Karnataka, India

S Amarappa
Dept of TCE, JNNCE
Karnataka, India

S V Sathyanarayana
Dept of ECE, JNNCE, Shimoga
Karnataka, India

Abstract - Speech is the most natural and effective method of communication between human beings. Automatic machine recognition of speech is a challenging task and it is essential for many applications in real life. In this context, we propose a novel model for spoken digit recognition specific to Kannada language. The proposed model is designed and simulated using a MATLAB code. Here, Mel Frequency Cepstral Coefficients (MFCC) features are extracted from speech signal and fed to support vector machine (SVM) classifier. SVM is trained using one-versus all approach with ten classifiers. The model is tested on various test samples and it is exhibiting satisfactory results with average recognition accuracy of 96.5%.

Keywords: Speech Recognition, Kannada Digit Recognition, Mel Frequency Cepstral Coefficients (MFCC), Support Vector Machine (SVM).

I. INTRODUCTION

Natural speech is the most effective method of communication. Automatic speech recognition is a process by which a speech signal is converted into equivalent text. Although speech is the most natural and effective method of communication, it is not easy to quickly review, retrieve, and reuse speech documents if they are simply recorded as audio signals. Therefore, transcribing speech is expected to become a crucial capability for the coming IT era [1]. A speech interface would support many valuable applications like telephone directory assistance, spoken database querying for novice users, "hands busy" applications in medicine, fieldwork, office dictation devices and for controlling electronic devices.

Isolated digit recognition technology opened up a class of applications called 'command-and-control' applications, in which the system is capable of recognizing a single word command from a small vocabulary and appropriately responding to the recognized command.[2]

Kannada is a language spoken in India predominantly in the state of Karnataka, making it the 33rd most spoken languages in the world. It is the official and administrative language of Karnataka state. Developing ASR for Kannada is interesting and challenging.

Classification is the task of choosing the correct class label for a given data input. Support vector machines (SVMs) are a set of new supervised learning methods used for binary classification. SVM is a structural risk minimization classifier algorithm derived from statistical learning theory by Vladimir Vapnik and his colleagues in 1992. Multiclass classification means a classification task with more than two classes. The dominant approach for multiclass classification is to reduce the single multiclass problem into multiple binary classification problems. Common approaches of reducing a multiclass problem into multiple binary classifiers include: (i) one-versus-the-rest also known as one-versus-all strategy and (ii) one versus- one approach [3]. The proposed model uses the first approach.

The rest of the paper is organized as follows. Section II discusses about literature survey. Design and implementation of the proposed model is discussed in section III. Results and discussions are included in section IV. Section V draws conclusions based on the results obtained.

II. LITERATURE SURVEY

The work on isolated digit recognition is carried in various languages across the globe and we mention here few of them: Plotkin E et al. (1977) [4] discusses about extracting information about speech segment, which relies on the binary structural representation of the signals reorganization isolated Hebrew digits. Rabiner L et al. (1979) [5] and Levinson S E et al. (1979) [6] demonstrated that clustering can be a powerful tool for selecting reference templates for speaker-independent word recognition. Sanches I et al. (1990) [7] worked on speaker-independent isolated digit recognition spoken in Portuguese. Maruti Limkara et al. (2012)[8] proposes an approach to recognize spoken English digits. Santosh V et al. (2013) [9] proposed an efficient speech recognition system for speaker-independent isolated digits using the Weighted MFCC (WMFCC). The experiments based on TI-Digits corpus. Elrgaby M et al. (2014) [10] proposes a scheme for recognizing isolated spoken Arabic digits, based on the Discrete Wavelet Transform (DWT) features. Azam S M et al. (2007) [11] implemented isolated spoken Urdu digits recognition using back propagation neural network.

Manaileng M J et al. (2013) [12] presents the development of a speech recognition system for automatically recognizing fluently spoken digit strings in Northern Sotho. The Cepstral Mean Vector Normalization (CMVN) increases the robustness of speech recognition systems.

Isolated digit recognition is employed on Indian languages also. Some of such works are discussed here:

Bhardwaj I et al. (2012) [13] discussed speaker dependent, multi speaker and speaker independent isolated words recognition in Hindi Language based on the HMM. Ghanty S K et al. (2010) [14] presents a method for recognizing isolated spoken Bengali numerals. MFCC have been used for feature and vector quantization is applied to reduce the dimension of the feature vectors. The classification is based on the dynamic time warping (DTW) and a minimum distance classifier based on Euclidean distance measure. Cini Kurian C et al. (2009) [15] and Renjith S et al. (2013) [16] presents speaker independent speech recognition system for Malayalam digits. The system employs MFCC as feature for signal processing and HMM for recognition. Cini Kurian et al. (2010) [2]have implemented speech recognition of Malayalam isolated digit by using Mel Frequency Cepstral Coefficients (MFCC) and Support Vector Machines. (SVM). Muralikrishna H et al. (2013) [17] and Shashidhara Nimbargi et al. (2015) [18] implemented a Kannada isolated digit recognition system using MFCC as feature vector and HMM as pattern recognizer

Based on the literature survey, it is seen that only few work is being taken up in isolated kannada digit recognitions [17-18]. In our previous work [19] we have developed a model for isolated Kannada digit recognition using Artificial Neural Network. However lot of work is still to be carried out. This motivated us to take up this experimental work.

III. DESIGN AND IMPLEMENTATION OF PROPOSED MODEL

In this paper, we propose a model to recognize spoken Kannada digits. The recorded speech samples are preprocessed and MFCC features are extracted. Then the extracted features are given to SVM classifier to classify the sample. Following sections deals with the details of various stages of the implementation.

A. PREPROCESSING

Speech is a highly non-stationary signal and hence speech analysis must be carried out on short segments, across which the speech signal is assumed to be stationary. With this view, the speech signal is divided into frames of small durations, typically 10 to 30ms with an overlap of 5 to 15ms. Here the frames of 10 ms consisting 80 samples without overlapping are considered.

Let x[n] be a speech signal with a sampling frequency of fs, and is divided into P frames each of length N samples such that $\{\vec{x}_1(n), \vec{x}_2(n), \vec{x}_3(n).....\vec{x}_P(n)\}$, where $\vec{x}_i(n)$ denotes the ith frame of the speech signal x[n] and is given by

$$\vec{x}_i(n) = \{x[i*N+n]\}_{n=0}^{N-1} \quad (1)$$

The speech signal x[n] is represented in a matrix notation of size N x P, where N=80 and P=100.

The problem of locating the beginning and end of a speech utterance in a background of noise is of importance. The selection of speech signal that correspond to a speech will eliminate the significant computation. The voiced part is extracted based on two simple time domain measurements: energy and zero crossing rates [20]. The amplitude of the speech signal varies appreciably with time and the amplitude of unvoiced segments is generally much lower than the amplitude of voiced segments. The short-time energy of the speech signal provides a convenient representation that reflects these amplitude variations. The short-time energy is calculated using rectangular window as given in equation (2).

$$En = \sum_{m=-\infty}^{\infty} [x(m)\, w(n-m)]^2 \quad \text{; Where w(n) is a window function} \quad (2)$$

The model for speech production suggests that the energy of voiced speech is concentrated below 3 kHz because of the spectrum falloff introduced by the glottal wave, whereas for unvoiced speech, most of the energy is found at highest frequencies. In general, if the zero-crossing rate is high, the speech signal is unvoiced else it is voiced [21].

As high frequency components of speech have lesser amplitude, Pre-emphasis improves its SNR. The transfer function of the pre-emphasis filter is given in equation (3):

$$H(z) = 1 - az^{-1} \quad 0.9 \le a \le 1.0 \quad (3)$$

Where 'a' is the filter coefficient and is chosen as 0.9375. The output signal y(n) after pre-emphasis is given in equation (4) [22]:

$$y(n) = x(n) - a * x(n-1) \quad . \quad (4)$$

B. MFCC FEATURE EXTRACTION

The Mel-Frequency Cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. It is a process of extracting features from the input signal by reducing the dimension of the input-vector still maintaining the uniqueness of the signal. The outline of the computation of MFCC is shown in Fig. 1.

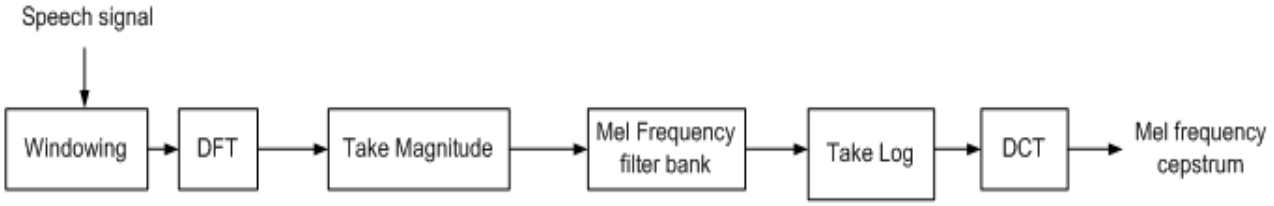

Fig. 1 Computation of Mel Frequency Cepstral Coefficients (features)

The MFCC features are extracted as follows:

Windowing: Multiply pth frame $\bar{x}_p(n)$ is with a hamming window function given in equation (5)

$$w(n) = \begin{cases} 0.54\text{-}0.46\cos\left(2\pi n / N\right) & \text{for } 0 \le n \le N\text{-}1 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Calculate Discrete Fourier Transform(DFT) of each frame to get spectrum.

Calculate the modulus of Fourier transform |X|.

|X| is warped according to the Mel scale using Mel frequency filter banks as given below: [23].

For any given frequency f, measured in Hz, Mel is calculated by the equation (6)

$$Mel(f) = 2595 * \log_{10}\left(1 + \frac{f}{700}\right) \quad (6)$$

|X| is segmented into a number of critical bands by means of a Mel filter bank which typically consists of a series of overlapping triangular filters defined by their center frequencies.

The parameters that define a Mel filter bank are number of Mel filte (F), minimum frequency (fmin) and maximum frequency (fmax.)

For speech, in general, it is suggested in [9] that fmin> 100 Hz. Furthermore, by setting fmin above 50-60Hz, we get rid of the hum resulting from the AC power, if present. It is known that, there is no much information above 6.8 KHz in human speech hence fmax be less than the Nyquist frequency.

The logarithm of the filter bank outputs is taken.

Finally, Discrete Cosine Transform(DCT) is taken to get MFCC features. Here each frame size is a vector of length 20.

C. CLASSIFICATION BY USING SUPPORT VECTOR MACHINE

Support vector machines (SVMs) are a set of new supervised learning methods used for binary classification. Given some data points, each belonging to one of two classes and the goal is to decide to which class a new data point belong. In support vector machines, a data point is viewed as an n-dimensional vector, in n-dimensional space Rn and we want to know whether we can separate such points with an (n − 1) dimensional hyper plane (Canonical plane). This is called a Linear Classifier. There are many hyperplanes that might classify the data. One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes, since in general the larger the margin the lower the generalization error of the classifier. The hyper plane is found by using the support vectors and margins. To calculate the margin, two parallel supporting hyper planes are constructed, one on each side of the Canonical plane, which is "pushed up against" the two data sets. So we choose the hyperplane such that the distance from it to the nearest data point on each side is maximized. If such a hyperplane exists, it is known as the maximum-margin hyperplane and the linear classifier it defines is known as a maximum margin classifier; or equivalently, the perceptron of optimal stability.[3]

SVM is a Machine Learning technique of classification and is a two-class classifier based on the use of Linear Discriminant Function g (x) = wT x + b, which represents a hyper plane in the feature space. It represents a straight line in two dimensional space, a plane in three dimensional space and an n-1 dimensional hyper plane in n dimensional space.

Multiclass SVM:

In digit recognition problem, there are ten classes (N=10) and ten SVM classifiers are constructed such as [m0, m1, m2, m3, m4, m5, m6, m7, m8, m9].

The training set for a classifier is: {(Xi,yi) }; where i=1,2…n} and n is the number of samples. Xi ϵ Rm , where m is the dimension of the feature vectors i.e., Xi=[xi1,xi2,…..xim] and yi {+1,-1}.

After training SVM mi, unknown feature vector 'p' is tested by using hyper plane described by wiT x + bi and classified as follows:

If $wi^T p + b_i \ge +1$ then feature vector 'p' belongs to class ci

Else if $wi^T p + b_i \le -1$ then feature vector p does not belongs to class ci.

The SVM 'm0' is trained such that it assigns positive group for class co and negative group for remaining classes. In general SVM 'mi' is trained to give positive result for class ci and negative result for rest of the classes. During testing phase a test sample is given to SVM 'm0', and if it does not belongs to class c0 then it is given to SVM 'm1' and the process is repeated till test sample is classified. The Pseudo code for SVM classifier is given in algorithm 1:

Algorithm 1: Pseudo code for classification

```
If (Svmclassify(m0,sample)) == 1 then
Sample is Digit 0
Elseif Svmclassify(m1,sample))==1 then
Sample is Digit 1
Elseif Svmclassify(m2,sample))==1 then
Sample is Digit 2
Elseif Svmclassify(m3,sample))==1 then
Sample is Digit 3
Elseif Svmclassify(m4,sample))==1 then
Sample is Digit 4
Elseif Svmclassify(m5,sample))==1 then
Sample is Digit 5
Elseif Svmclassify(m6,sample))==1 then
Sample is Digit 6
Elseif Svmclassify(m7,sample))==1 then
Sample is Digit 7
Elseif Svmclassify(m8,sample))==1 then
Sample is Digit 8
Elseif
Svmclassify(m9,sample))==1 then
Sample is Digit 9
Else
Sample is invalid
End
```

Each of the N classifiers is trained using all available samples. The classifications of digits using multi class SVM is shown in Figure 2. Here values ≥ 1 are normalized to +1 and values ≤ -1 are normalized to-1.

D. DATABASE CREATION:

In order to facilitate the training and testing of the recognizer, speech database is required. A variety of speech samples were obtained from different speakers to form the speech database. The collected database includes 100 speech samples from 10 different speakers aged between 20 to 35 years. 70% of the collected data is used for training and remaining is used for testing. Apart from this, a small data base is created for testing. The digits must be spoken clearly so that it avoids general variations and confusions.

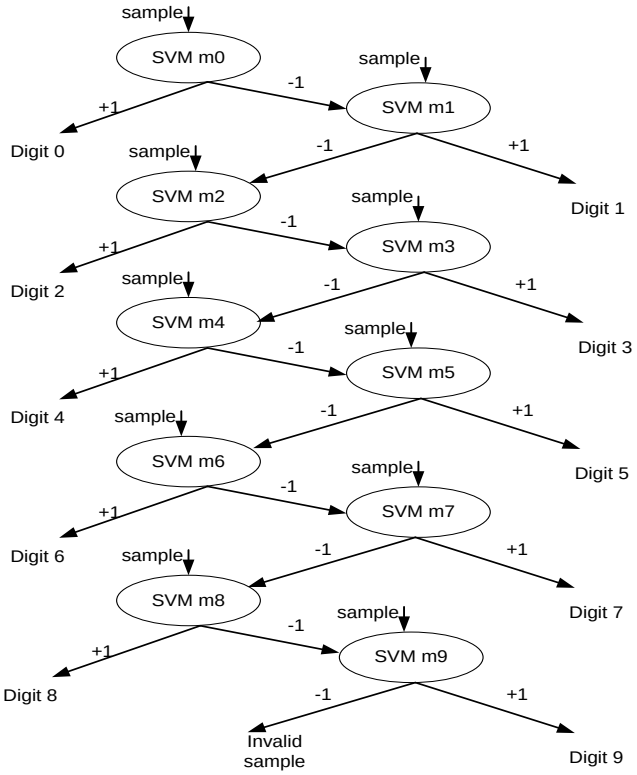

Fig. 2 multi class SVMs for digit classifications

IV. RESULTS AND DISCUSSION

The digit database consists of ten samples from "zero" to "nine" collected from 5 male speakers and 5 female speakers. The symbols of digits in Kannada are shown in Table 1 for each English digit. The database is divided into training and testing. Training set is used to train support vector machine.

Table 1 List of Kannada Digts

PÀ£ÀßqÀ ¸ÀASÉåUÀ¼ÀÄ (kannada digits in words)	¸ÉÆ£ÉÉß	MAzÀÄ	JgÀqÀÄ	ªÀÄÆgÀÄ	£Á®ÄÌ	LzÀÄ	DgÀÄ	K¼ÀÄ	JAlÄ	MA§vÀÄÛ
kannada digits	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
digit	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Testing set is used to test the performance of SVM. The trained SVM is tested for test samples of each digit collected from speakers. This recognizer can be very well adapted for voice dialing. The tested results are shown in Table 2.

Table II Confusion Matrix

	0	1	2	3	4	5	6	7	8	9
0	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%
1	0%	95%	0%	0%	0%	5%	0%	0%	0%	0%
2	0%	0%	95%	3%	0%	0%	0%	0%	0%	2%
3	0%	0%	0%	93%	0%	0%	7%	0%	0%	0%
4	0%	0%	0%	0%	98%	0%	0%	2%	0%	0%
5	0%	3%	0%	0%	0%	95%	0%	0%	0%	2%
6	2%	0%	0%	3%	0%	0%	95%	0%	0%	0%
7	0%	0%	0%	0%	0%	0%	0%	98%	2%	0%
8	0%	0%	0%	0%	0%	0%	0%	2%	98%	0%
9	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%

The digits 0 and 9 are recognized with an accuracy of 100% and there is no confusion with other digits. The digit 3 is recognized with an accuracy of 93% and confused with digit six 7%. The system developed works with an average accuracy of 96.5%, when it is tested with the test database. When it works in normal office environment, with a different group of people, the system accuracy reduces. In Kannada language all the digits have different acoustic characteristics and the error or less recognition accuracy is due to varying pronunciation of different speakers.

The waveform of the recorded speech for Kannada digit 9 is shown in Figure 3 a), which has 8000 samples. The energy of the signal is shown in figure 5 b). The zero crossing rate of the signal is shown in figure 5 c). based on energy and zero crossing rate the selected voice part of the recorded speech is shown in figure d).

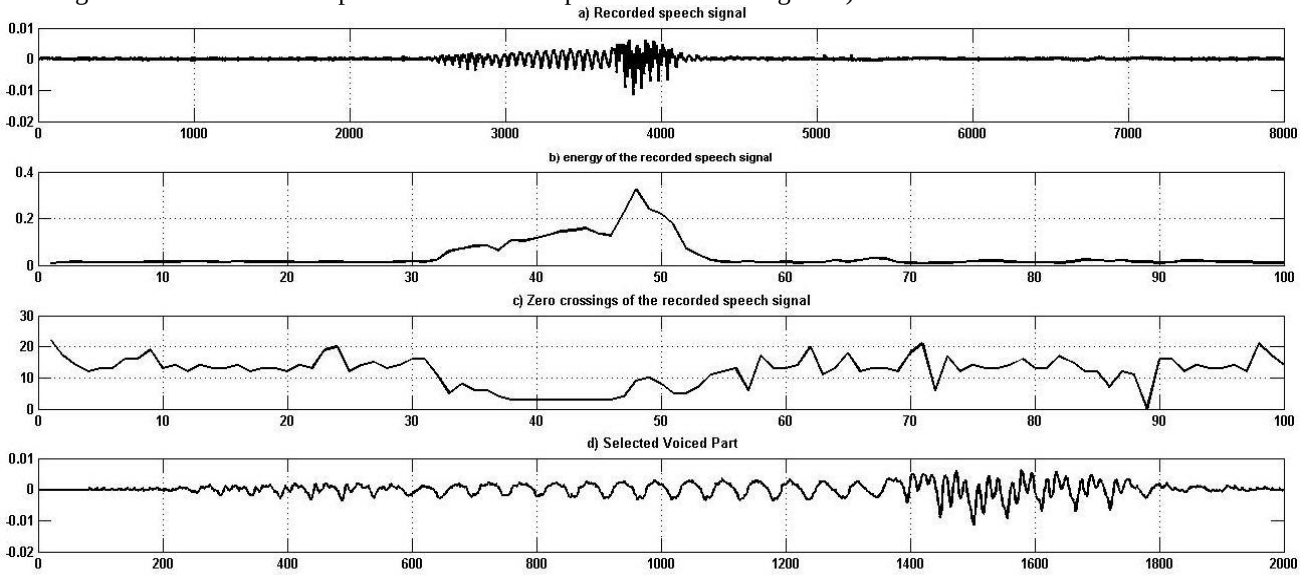

Fig. 3 Speech Waveform of a Kannada digit 9

V. CONCLUSION

In this paper, we have developed a model for Kannada isolated digit recognition using MFCC as feature vector and Multi class SVM as classifier. The proposed system is tested for 5 male and 5 female speech samples. It is found that the

system works with an average accuracy of 96.5%. However, this work is under progress and in future database with large and real time input samples will be tested and results will be analyzed to arrive at better conclusion. At this stage SVM appears to be a successful and powerful approach for effective classification. Digit recognition finds applications in voice dialing systems and it needs to be a speaker independent system. This system acts as a basis for real time speech recognition products, where input will never be a single word.

REFERENCES

[1] Sadaoki Furui, Tomonori Kikuchi, Yousuke Shinnaka, and Chiori Hori *Speech-to-Text and Speech-to-Speech Summarization of Spontaneous Speech*, IEEE Transactions On Speech And Audio Processing, VOL. 12, NO. 4, JULY 2004, pp 401- 408.

[2] Renjith S. Aju Joseph Anish Babu K.K." *Isolated Digit Recognition for Malayalam- An Application Perspective*", International Conference on Control Communication and Computing (ICCC), 2013, Pp 190- 193.

[3] S Amarappa, Dr. S V Sathyanarayana "*Data classification using Support vector Machine (SVM), a simplified approach*", International Journal of Electronics and Computer Science Engineering,Volume 3, Number 4,2014, ISSN 2277-1956, PP 435-445.

[4] Plotkin, E. Plotkin, N. Polevoi, Y." *Recognition of spoken digits by joint segmentation of envelopes of two-signal transforms*", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '77. 1977, (Volume:2), Page(s): 456 – 459.

[5] Rabiner, L, Wilpon, J.G., "*Speaker-independent isolated word recognition for a moderate size(54 word)vocabulary***",** IEEE Transactions on Acoustics, Speech and Signal Processing, Year: 1979, Volume: 27, Issue: 6, Pages: 583 – 587.

[6] Levinson S.E, Rabiner L, Rosenberg A.E, Wilpon J.G.," *Interactive clustering techniques for selecting speaker-independent reference templates for isolated word recognition***",** IEEE Transactions on Acoustics, Speech and Signal Processing, Year: 1979, Volume: 27, Issue: 2, Pages: 134 – 141.

[7] Sanches I, Alens N.," *A speaker-independent digit recognizer***",** Telecommunications Symposium, 1990. ITS '90 Symposium Record., SBT/IEEE International Year: 1990, Pages: 202 – 206.

[8] MarutiLimkara, RamaRaob, VidyaSagvekarc," *Isolated Digit Recognition Using MFCC AND DTW*", International Journal on Advanced Electrical and Electronics Engineering, (IJAEEE), ISSN (Print): 2278-8948, Volume-1, Issue-1, 2012, Pp 559-64.

[9] Santosh V. Chapaneri, Dr. Deepak J. Jayaswal "*Efficient Speech Recognition System for Isolated Digits*", International Journal of Computer Science & Engineering Technology ISSN : 2229-3345 Vol. 4 No. 03 Mar 2013, Pp 228-236.

[10] Elrgaby M, Amoura A, Ganoun A.," *Spoken Arabic Digits Recognition Using Discrete Wavelet***",** 16th International Conference on Computer Modelling and Simulation (UKSim), 2014 UKSim-AMSS Year: 2014, Pages: 275 – 279.

[11] Azam S.M, Mansoor Z.A, Mughal M.S, Mohsin, S.,"*Urdu Spoken Digits Recognition Using Classified MFCC and Backpropgation Neural Network***",** Computer Graphics, Imaging and Visualisation, 2007. CGIV '07 Year: 2007, Pages: 414 – 418.

[12] Manaileng M.J. Manamela M.J, "*Connected-digits recognition for an under-resourced language using Hidden Markov Models*", ELMAR, 2013 55th International Symposium, 25-27 Sept. 2013, ISSN :1334-2630 Page(s):211 – 214.

[13] Bhardwaj I, Londhe N.D.**,"** *Hidden Markov Model based isolated Hindi word recognition***",** , 2nd International Conference on Power, Control and Embedded Systems (ICPCES), 17-19 Dec. 2012 ,Page(s):1 – 6

[14] Ghanty S.K, Shaikh S.H, Chaki, N.," *On recognition of spoken Bengali numerals*", International Conference on Computer Information Systems and Industrial Management Applications (CISIM), Year: 2010 , Pages: 54 – 59.

[15] Cini Kurian, Kannan Balakrishnan, "*Speech Recognition of Malayalam Numbers*", World Congress on Nature & Biologically Inspired Computing (NaBIC 2009), PP 1475-1479.

[16] Cini Kurian, Firoz Shah.A, Kannan Balakrishnan, "*Isolated Malayalam Digit Recogntion Using Support Vector Machines*", , IEEE International Conference on Communication Control and Computing Technologies (ICCCCT), 2010 PP 692-695.

[17] Muralikrishna H, Ananthakrishna T, Dr. Kumara shama, "*HMM Based Isolated Kannada Digit Recognition System using MFCC*",
International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013, Pp 730-733.

[18] Shashidhara Nimbargi, Dr.S.N. Chandrashekara, "*Isolated Speaker Independent Kannada ASR System Using HTK*", International Journal of Combined Research & Development (IJCRD) eISSN: 2321-225X;pISSN: 2321-2241 Volume: 4; Issue: 6; June -2015, PP – 650 – 653.

[19] Harisha S B, Amarappa S, Dr. S V Satyanarayana " *Automatic Speech Recognition – A Literature Survey on Indian Languages and Ground Work for Isolated Kannada Digit Recognition using MFCC and ANN*",

International Journal of Electronics and Computer Science Engineering, ISSN 2277-1956/V4 N1, 99. February 2015, PP 91-105.

[20] L.R Rabiner, M.R. Sambur, "*An Algorithm for Determining the Endpoints of Isolated Utterances*", Bell Syst. Yech. J., Vol 54, No. 2, pp297-315, February 1975.

[21] L.R Rabiner, R.W Schafer: "*Digital Processing of Speech Signal*", Pearson Education (Singapore), 2005.

[22] M. H. Moattar, M. M. Homayounpour, "*A simple but efficient real-time Voice Activity Detection algorithm*", 17th European Signal Processing Conference (EUSIPCO), pp. 2549 – 2553. , Aug. 2009

[23] S. Molau, M. Pitz, R. S. Uter, and H. Ney, "*Computing Mel-frequency cepstral coefficients on the power spectrum*," Proc. Int. Conf. on Acoustic, Speech and Signal Processing, pp. 73 – 76, 2001.

