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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Kannada Speech Recognition Using MFCC and KNN Classifier for Banking Applications

S B Harisha, S Amarappa, S V Sathyanarayana

Assistant Professor, Dept. of TCE, JNNCE, Shimoga, India

Associate Professor, Dept. of TCE, JNNCE, Shimoga, India

Professor, Dept. of ECE, JNNCE, Shimoga, India

ABSTRACT: This paper presents a Kannada speech recognition system. Isolated speech recognition includes speech segmentation, feature extraction and classification. In our work, a blind speech segmentation procedure is being used to segment the spoken Kannada sentences into words using the endpoint detection technique. MFCC signal analysis technique is used to extract the features of segmented words. K- Nearest Neighbor (KNN) classifier is used as a classifier. "Filling a withdrawal slip of a bank" is considered as an application to test the performance of this system. The proposed system is designed and simulated using MATLAB. The system developed has achieved the recognition accuracy of about 91.5% when it is tested for the small vocabulary environment.

KEYWORDS: MFCC feature, KNN classifier, Energy, Spectral flux, magnitude.

I. INTRODUCTION

A word-based (WB) approach to speech recognition is popular for its simplicity in implementation and for its good performance for small-to-medium size vocabulary, isolated word recognition tasks. In an Automatic Speech Recognition (ASR) system, the vocabulary is made up of all the words that it can recognize. An ASR system that can recognize a small number of words (say, 1 to 100) is called as small vocabulary system. The size of vocabulary of a speech recognition system affects the complexity, processing requirements and the accuracy of the system. Some applications require only a few words (e.g. numbers only); others require very large dictionaries (e.g. dictation machines).

Manual segmentation can be used but it has two major drawbacks: i) The process is both laborious and tedious, requiring extensive listening and spectrogram interpretation. ii) Due to the subjective nature of a manual segmentation, there will be inconsistencies from trial to trial, even for segmenting the same utterance. In order to alleviate these problems, automatic procedures for segmenting speech into word units are needed [1].Automatic speech segmentation can be classified into two types: Blind segmentation and Aided segmentation algorithms. In blind segmentation there is no use of pre existing or external knowledge of linguistic properties. The aided segmentation uses some sort of external linguistic knowledge of the speech. Generally there are two kinds of segmentation: Phonemic segmentation and syllable like unit segmentation. Phonemic segmentation segments speech sequence into small phonemes and syllable segmentation segments speech into syllables [2].

Kannada is an Indian language predominantly used in the state of Karnataka and it is the 33rd largest spoken languages in the world. It is the official and administrative language of Karnataka state. Hence, Developing ASR for Kannada is interesting and challenging.

Classification is an area of machine learning that takes raw data and classifies it as belonging to a particular class based on the required parameter set. The belief inherited in Nearest Neighbor Classification is quite simple; test samples are classified based on the class of their nearest neighbors. For example, if it walks like a duck, quacks like a duck, and looks like a duck, then it's probably a duck.

In this method if a data "x" has k nearest data where majority of them are having the same label "y", then "x" belongs to "y". The Euclidian distance can be calculated using equation (1) . If two vectors x_i and x_j are given where x_i $=(x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}\ldots. x_{in})$ and $x_j =(x_{j1}, x_{j2}, x_{j3}, x_{j4}, x_{j5}\ldots. x_{jn})$ The difference [3] between xi and x_j is

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

$$D(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2} \qquad (1)$$

The advantages of the KNN classifier are 1) Simple and easy to learn. 2) Robust to noisy training data and 2) Effective if training data is large. The disadvantages are 1) Biased by value of k. 2) Computation Complexity and 3) Being a supervised learning lazy algorithm i.e. runs slowly.

In this work, KNN classifier is used just as a random choice as it is quite simple. However, other classifiers can also be used.

A withdrawal slip is a bank document on which a person writes the date, account number and amount of money to withdraw from a bank. It is called a withdrawal slip because it is used to make a withdrawal from a person's account. It includes important information that allows the bank to keep an accurate record of the withdrawal and provide the required amount.

In this paper, we have developed a small vocabulary Kannada speech recognition system consisting of eighteen words. The vocabulary is selected such that it is suitable for application like filling a withdrawal slip in a bank, which may be useful to illiterate customers of bank.

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

In the early 1920s machine recognition came into existence. The first machine to recognize speech to any significant degree was commercially named, Radio Rex (toy, manufactured in 1920) [4]. Research in speech technology began in early 1936 at Bell Labs. In 1939, Bell Labs demonstrated a speech synthesis machine (which simulates talking) at the World Fair in New York. The earliest attempts to devise systems for ASR by machine were made in 1950s, when various researchers tried to exploit the fundamental ideas of acoustic phonetics.

Ang F et al. (2015) [5] proposed the use of cepstral features derived from time-varying linear predictive coding, where the autoregressive model of the speech signal is represented by coefficients that are linear combinations of some simple basis functions for a 142 vocabulary, isolated words Japanese speech recognition task. Principi E et al. (2014) [6] presents power normalized cepstral coefficients based super vectors and i-vectors for small vocabulary speech recognition. Experimental results showed the appropriateness of the super vector and i-vector based solutions with respect to the other state-of-the art techniques here addressed. Ren Wenxia et al.(2009) [7] developed small vocabulary, isolated word speech recognition system. In software the endpoint detection and the adaptive speech recognition arithmetic were used. The speech recognition rate rises up to 96%. Lucey S et al.(2001) [8] presents a improving visual noise insensitivity in small vocabulary audio visual speech recognition applications. The use of a high dimensional secondary classifier on the word likelihood scores from both the audio and video modalities is investigated for the purposes of adaptive fusion. Preliminary results are presented demonstrating performance above the catastrophic fusion boundary for our confidence measure irrespective of the type of visual noise presented to it. Chengalvarayan R (1997) [9] presents adaptation of quadratic trajectory segment models for small vocabulary speech recognition. They have implemented a speech recognizer using TI46 corpora. Experimental results show that the quadratic trended HMM always outperforms the standard, stationary-state HMM and that adaptation of quadratic polynomial coefficients only is better than adapting both polynomial coefficients and precision matrices when fewer than four adaptation tokens are used. Kamm C A et al. (1994) [10] discussed the Speech recognition issues for directory assistance applications. Speech recognition performance for a set of 200 spoken names was measured for directories ranging from 200 to 1.5 million unique names. Recognition accuracy decreased from 82.5 percent for a 200-name directory to 18.5 percent for a 1.5 million name directory. Velez E et al.(1991) [11] describe first phase in the development of a speech recognition system for small vocabularies. The system is designed to handle speaker-independent continuous speech and can be easily modified for different vocabularies. The system consists of a spectral analysis stage followed by vector quantization (VQ) and hidden Markov modeling (HMM). Preliminary experiments

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

on continuous phoneme and digit recognition were performed on an unrestricted-speaker telephone database. Walker K et al.(1989) [12] presented a speaker independent automatic speech recognition system for a small vocabulary, employing phonetically based methods. The system uses formant tracking and relative energy values to characterize each word in the vocabulary (the digits, 0 to 9) The system was tested on a number of speakers of both sexes, with encouraging results. Brognaux S et al. (2016) [13] implemented a HMM-Based Speech Segmentation. The obvious advantage of this technique is that it is applicable to any language or speaking style and does not require manually aligned data. Receveur S et al.(2016) [14] successfully applied the turbo principle to the domain of ASR and thereby provide solutions to the above mentioned information fusion problem. On a small vocabulary task, their proposed turbo ASR approach outperforms even the best reference system on average over all SNR conditions and investigated noise types by a relative word error rate (WER) reduction of 22.4% (audio-visual task) and 18.2% (audio-only task), respectively

Anusuya et al. (2010) [15] have designed Isolated Words Recognizer for Kannada Language speech, based on the Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA). First, the DWT of the speech is computed and then MFCC coefficients are calculated. For this, PCA procedure is applied for speech recognition. Here they create the database of Kannada isolated digits from 0 to 10. Hemakumar G et al. (2014)[16] designed algorithm recognizes spoken Kannada words independent of speakers. The proposed method normalizes the original speech signal of every isolated word and extracts Linear-Predictive coding (LPC) coefficients, and converts them into Real Cepstrum Coefficient. For experimentation, they have used 294 unique Kannada words. The success rate of the proposed system for known speaker data is 98.29 % and unknown speaker data is 91.66 %. The accuracy rate of individual word can increase by using more number of training sets and this can also lead to increase in the number of word recognition in the first hit.

Renjith S. et al. (2013)[17]presented speaker independent speech recognition system for Malayalam digits. The system employs MFCC as feature for signal processing and HMM for recognition. Raji Sukumar, et al. (2010) [18] discussed a novel technique for recognition of the isolated question words from Malayalam speech query. They have created and analyzed a database consisting of 500 isolated question words. Fast Fourier Transform (FFT) and Discrete Cosine transform (DCT) is used for the feature extraction purpose and ANN is used for classification and recognition. P. Punitha et al. (2014) [19] described the design of an algorithm to recognize continuous Kannada speech using HMM Method in the speaker dependent mode. LPC coefficients are used as features. K-means procedure is performed on the feature vectors to obtain the observation sequence. Muralikrishna, H. et al. (2013) [20] has implemented Kannada isolated digit recognition system using MFCC as feature vector and HMM as pattern recognizer. Performance of the system is evaluated and compared based on the MFCC along with its first and second order derivatives. Harisha S B et al. (2015) [21] proposed a novel model for spoken digit recognition specific to Kannada language. The proposed model is designed and simulated using a MATLAB code MFCC features are extracted from speech signal and fed to support vector machine (SVM) classifier.

In English and other foreign languages, lots of work has been found in this field. However, a good amount of work has been done in speech recognition with respect to Hindi, Punjabi, Tamil, Telugu, Bengali and Marathi Languages. In Kannada language ASR work is not explored as much as other major spoken Indian languages [22]. This motivated us to take up ASR in Kannada as the research topic.

III. DESIGN AND IMPLEMENTATION OF PROPOSED MODEL

In this paper, we propose a model to recognize spoken Kannada words. The recorded speech samples are pre-processed and MFCC features are extracted. Then the extracted features are given to KNN classifier to classify the sample. Finally recognized words are written on a withdrawal slip image. Following sections deal with the details of various stages of the implementation.

A. PRE-PROCESSING

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.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

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 i[th] frame of the speech signal x[n] and is given by $\vec{x}_i(n)=\{x[i*N+n]\}_{n=0}^{N-1}$ (2)

The speech signal x[n] is represented in a matrix notation of size N x P, where N=80 and P depends on the duration of the recorded speech sentence.

B. SEGMENTATION

Speech segmentation is used to segments continuous speech into uniquely identifiable or phonemes, syllables, words or sub words. Segmentation is an important role in speech recognition to reduce memory size and computational complexity. Segmentation is used to detect the proper start and end point of speech events. [2]

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 measurements (speech features) like energy, zero crossing rates, spectral flux and spectral centroid and magnitude. Energy, magnitude and spectral flux are used in the proposed system.

(i) Short Time Energy

The energy associated with speech is time varying in nature. Hence the interest for any automatic processing of speech is to know how the energy is varying with time and to be more specific. By the nature of production, the speech signal consist of voiced, unvoiced and silence regions. Further the energy associated with voiced region is large compared to unvoiced region and silence region will not have least or negligible energy.

Thus short term energy can be used for voiced, unvoiced and silence classification of speech. The energy of a signal is typically calculated on a short- time basis, by windowing the signal at a particular time, squaring the samples and taking the average.

Short time energy is computed for each frame using

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

(ii) Spectral Flux

Spectral flux refers to a measure of how quickly the power spectrum of a signal is changing, calculated by comparing the power spectrum for one frame against the power spectrum from the previous frame [2]. The spectral flux is given by

$SF_i = \sum_{k=1}^{N/2} [|X_i(k)| - |X_i(k-1)|]^2$ Here $X_i(k)$ is the DFT coefficients of i[th] frame.

(iii) Magnitude

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. Each frame of the signal is passed through 9[th] order Butterworth low pass filter and the magnitude of the response of the filter is calculated as [23]

$\text{magnitude}=\sum_{i=1}^{80} |Y_i|$ Where Y_i is the response of the filter for i[th] frame.

Proposed Algorithm for Segmentation:

After computing the speech features, a simple algorithm is applied to detect the speech words segments as given in Algorithm 1:

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Algorithm 1: Speech words segmentation

1. Calculate the threshold: threshold=mean(energy)/2;
2. Find the frames whose energy is greater than the threshold. In the rest of the discussion these set of frames are termed as energy frames.
3. First value in the energy frame is the beginning point of first word.
4. If the difference of two successive frames is greater than 18, then first among these two is the end point of first word and the latter one is beginning point of next word. This procedure is repeated till the end of the energy frames. The value 18 is fixed by trial and error method through experimentation.
5. Length of each word Li and the distance between successive word dij is calculated.
6. If the length Li is less than minimum word length and distance dij is greater than minimum space between two words then that word is discarded.

The above process is applied for both magnitude and spectral flux and number of words are calculated, it should be noted that, the rest of the implementation will proceed if and only if the number of words calculated from each of these feature i.e., energy, spectral flux and magnitude are same.

The beginning of each word is taken such that it will be the minimum index of each word calculated by considering each feature. For the end point of the frame, maximum index of each word is considered.For each word 40 frames are selected, depending on the final end point calculation of each word they are either stretched or compressed from both the points.

For each segmented word, MFCC features are extracted which is explained in the following section.

C. Pre emphasis

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):

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

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) [24]:

$$y(n) = x(n) - a * x(n-1) \tag{4}$$

D. 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.

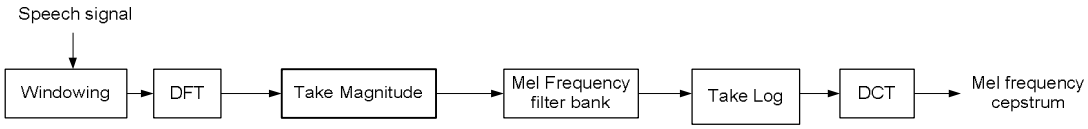

Fig. 1 Computation of Mel Frequency Cepstral Coefficients (features)

The MFCC features are extracted as given in algorithm 2:

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

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

(ii) DFT: calculate DFT of each frame to get spectrum.

(iii) Magnitude: Calculate the modulus of Fourier transform |X|.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

(iv) Mel frequency filter banks: |X| is warped according to the Mel scale [25].

- 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 (a) number of Mel filters, F (b) minimum frequency, fmin and (c) maximum frequency, fmax.
- For speech, in general, it is suggested in [24] 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.

(v) Apply logarithm: The logarithm of the filter bank outputs is taken.

(vi) DCT: Finally, DCT is taken to extract MFCC features. Here each frame size is a vector of length 13.

E. K-NEAREST NEIGHBOR CLASSIFIER

Among the various methods of supervised statistical pattern recognition, the Nearest Neighbor rule achieves consistently high performance, without a priori assumptions about the distributions from which the training examples are drawn. A new sample is classified by calculating the distance to the nearest training case. The k-NN classifier extends this idea by taking the k nearest points and assigning the class of the majority. It can be shown that the *k* – nearest neighbor rule becomes the Bayes optimal decision rule as k goes to infinity [26]. However, it is only in the limit as the number of training samples goes to infinity that the nearly optimal behavior of the *k* - nearest neighbor rule is assured.

In K-Nearest Neighbor method, the selection of k values is tricky and application dependent. To simplify our problem it is always fixed to odd number (typically 1, 3 or 5) so that no tie can happen. Larger k values help reduce the effects of noisy points within the training data set, and the choice of k is often performed through cross-validation. In our experiment we tried to classify the words data set for different values of k and for k=5 we got the maximum correct classification rate.

K – Nearest neighbor decision rule:
Let (x_i,c_i) ; i=1,2,….n be given, where $x_i \in R^m$; i=1,2,…n; and ci denote the label of x_i for each i.
Let us assume that the number of class is C, C≥2. i.e., $c_i \in \{1,2,\ldots\ldots C\}$ for all i.
Let x be a point for which the label (class) is not known. We need to find the label (class) of x. the following steps gives the general KNN algorithm

1. Let k be a positive integer
2. Calculate the distance d(x, x_i) for all i=1,2,3….n where ‘d’ denotes the Euclidean distance
3. Arrange the ‘n’ distance in non decreasing order
4. Take the first k distances
5. Find those k points corresponding to these k-distances.
6. Let K_i denote the number of points belonging to the i^{th} class among the k-points; i=1,2, ..C (k_i for some class can be zero and sum of all K_i is equal to k)
7. Put x in class i if $K_i > K_j$ for all $j \neq i$

Proposed system using KNN Algorithm:
We have C=18 words (classes) and each word is recorded by 25 speakers. So, training data consist of n=450 sample. Sample data is classified using Algorithm 32

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Algorithm 2: KNN Classifier

1. In our experiment we tried to classify the words data set for different values of k and for k=5 we got the maximum correct classification rate.
2. Calculate the "n" Euclidean distance between sample data and each of the training data.
3. Arrange the n distances in increasing order.
4. Select the first k = 5 distances in the sorted distances.
5. Find the words corresponding to these k distances
6. Let K_i denote the number of words belonging to the i^{th} class among the k-points; i=1,2, ...18
7. Put the sample data to word or class i if $K_i > K_j$ for all $j \neq i$

F. DATABASE CREATION

In order to facilitate the 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 450 speech samples from 25 different speakers aged between 20 to 35 years. The digits must be spoken clearly so that it avoids general variations and confusions. The speech is recorded using PRAAT software and a 16-20KHz microphone.

The vocabulary consist of 18 words as shown in table 1 in which there are 2 names, 10 digits and 6 labels.

Table 1: list of vocabulary words

Sl No.	Word Type	Word	Word in Kannada	Sl No.	Word Type	Word	Word in Kannada
1.	Name	Harisha	ºÀjÃ±À	10.	Digit	One	MAzÀÄ
2.	Name	Amarappa	CªÀÄgÀ¥Àà	11.	Digit	Two	JgÀqÀÄ
3.	Label	Name	ºÉ¸ÀgÀÄ	12.	Digit	Three	ªÀÄÆgÀÄ
4.	Label	Account	SÁvÉ	13.	Digit	Four	£Á®ÄÌ
5.	Label	Amount	ªÉÆvÀÛ	14.	Digit	Five	LzÀÄ
6.	Label	Thousand	¸Á«gÀ	15.	Digit	Six	DgÀÄ
7.	Label	Hundred	£ÀÆgÀÄ	16.	Digit	Seven	K¼ÀÄ
8.	Label	Fifty	LªÀvÀÄÛ	17.	Digit	Eight	JAlÄ
9.	Digit	Zero	¸ÉÆ£Éß	18.	Digit	nine	MA§vÀÄ

G. An Application: "Filling a Withdrawal Slip in a Bank"

The proposed system is tested for a real time application *"filling a withdrawal slip in a bank"* which may be very useful in day-to-day life especially for illiterate bank customers.

A withdrawal slip as shown in figure 2 is considered in this application. In withdrawal slip we have to fill the name, amount in words, account number, amount in digits and date. The sentence should be spoken in the order of name, account number and amount. Few sentences are shown in Table 2.

The two names considered are amarappa and harisha. The account number will be of 3 digits. The amount can be of at most 4 digits and at least 3 digits. The amount can be Rs. 4150, 4100 and can not be 4160 or 4120. The date can be read from the system and according to that it is updated. A sentence can have at most 12 words and at least 9 words depending on the amount as shown in Table 2:.

Table 2: possible words in a sentence

Word1	Word2	Word3	Word4	Word5	Word6	Word7	Word8	Word9	Word 10	Word11	Word12
Hesaru (Name)	Amarappa Harisha	Khathe (Account)	Digit 0to9	Digit 0to9	Digit 0to9	Mottha (Amount)	Digit 0to9	Savira nuru	Digit 0to9 aivatthu	Nuru	aivatthu

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

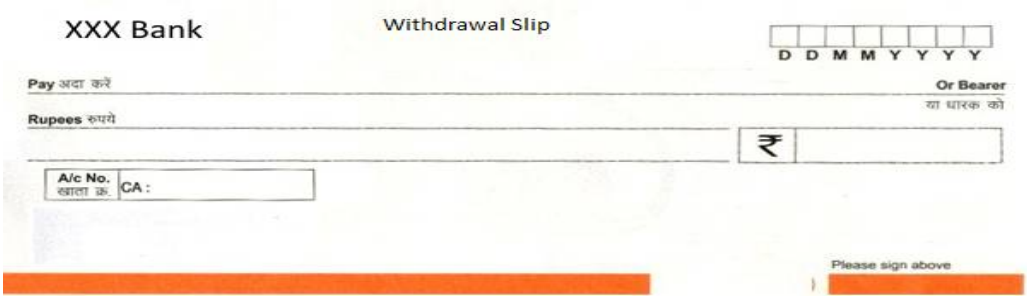

Fig. 2: blank withdrawal slip

IV. RESULTS AND DISCUSSION

As discussed in earlier section, an application that can be used in bank to withdraw the money for an illiterate bank customer is considered. Here, the customer can speak his requirement through microphone, which would be recognized by our system and the withdrawal form will be accordingly printed. About, 450 speech samples collected from 25 different speakers are considered for testing. It was observed that the system works efficiently for a small vocabulary environment.

Figure 3 shows the wave form of the speech signal for sentence 2 of Table 3 and the features considered for segmentations (energy, magnitude and spectral flux). The proposed system is tested for 10 sentences and five sentences are listed in Table 3

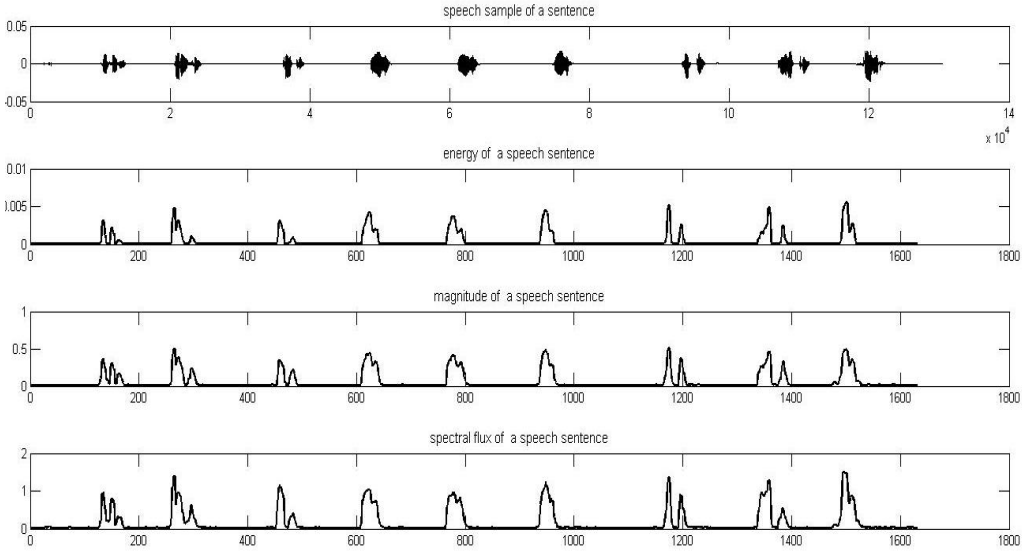

Fig. 3. a) speech wave form of a sentence Name Amarappa Account 143 amount 1 hundred b) energy c) magnitude and d) spectral flux

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Table 3: List of sentences used for testing

Sl No.	Sentence	No. of words
1.	ºÉ¸ÀgÀÄ CªÀÄgÀ¥Àà SÁvÉ MAzÀÄ JgÀqÀÄ ªÀÄÆgÀÄ ªÉÆvÀÛ MAzÀÄ ¸Á«gÀ JgÀqÀÄ £ÀÆgÀÄ Name Amarappa Account 123 amount 1 thousand 2 hundred	11
2.	ºÉ¸ÀgÀÄ CªÀÄgÀ¥Àà SÁvÉ MAzÀÄ £Á®ÄÌ ªÀÄÆgÀÄ ªÉÆvÀÛ MAzÀÄ £ÀÆgÀÄ Name Amarappa Account 143 amount 1 hundred	9
3.	ºÉ¸ÀgÀÄ ºÀjÃ±À SÁvÉ ¸ÉÆ£Éß LzÀÄ £Á®ÄÌ ªÉÆvÀÛ JAlÄ ¸Á«gÀ JgÀqÀÄ £ÀÆgÀÄ Name Harisha Account 054 amount 8 thousand 2 hundred	11
4.	ºÉ¸ÀgÀÄ ºÀjÃ±À SÁvÉ K¼ÀÄ K¼ÀÄ K¼ÀÄ ªÉÆvÀÛ £Á®ÄÌ ¸Á«gÀ Name Harisha Account 777 amount 4 thousand	9
5.	ºÉ¸ÀgÀÄ ºÀjÃ±À SÁvÉ DgÀÄ K¼ÀÄ JAlÄ ªÉÆvÀÛ LzÀÄ £ÀÆgÀÄ LªÀvÀÄÛ Name Harisha Account 678 amount 5 hundred fifty	10

The accuracy of word segmentation is 98.25% as given in Table 4. The accuracy will reduce slightly when the speech is noisier.

Table 4: Segmentation results

Sentence	Number of word segments expected	Number of words segmented by system	Segmentation accuracy %
1	11	11	100
2	9	9	100
3	11	10	90.9
4	9	9	100
5	10	10	100
6	12	12	100
7	10	10	100
8	11	11	100
9	12	11	91.66
10	9	9	100
Average segmentation rate			98.25

For isolated words average recognition accuracy is 96.2% as given in Table 5.

Table 5: Isolated word recognition accuracy

Sl No.	Word	Accuracy %	Sl No.	Word	Accuracy %
1.	Harisha	100	10.	One	92
2.	Amarappa	96	11.	Two	100
3.	Name	96	12.	Three	92
4.	Account	100	13.	Four	100
5.	Amount	100	14.	Five	96
6.	Thousand	96	15.	Six	92
7.	Hundred	96	16.	Seven	100
8.	Fifty	96	17.	Eight	92
9.	Zero	92	18.	nine	96

Out of 10 sentences 7 sentences are recognized correctly and as a number of words out of 104 words 95 words are recognized correctly. For sentences (connected words) recognition accuracy is 91.5% as given in Table 6.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Table 6: Sentences word recognition accuracy

Sentence	Number of word segments	Number of words segments recognized correctly	accuracy%
1	11	11	100
2	9	9	100
3	11	07	63.6
4	9	9	100
5	10	10	100
6	12	12	100
7	10	08	80
8	11	11	100
9	12	9	75
10	9	9	100
	104	95	
Average Accuracy in sentences			91.5

The recognition accuracy of words in sentences is less compared to isolated word. Figure 4 shows the image of a withdrawal slip with the recognized words for sentence 4.

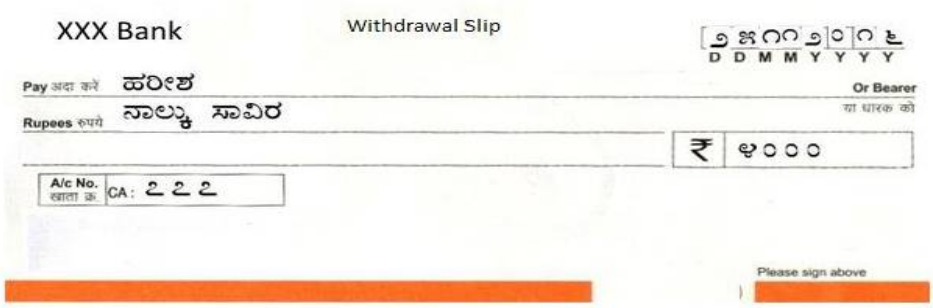

Fig. 4: Withdrawal slip with recognized spoken words

V. CONCLUSION

In this paper, we have developed a model for the small vocabulary Kannada speech recognition consisting of eighteen words. Segmentation of speech into word is accomplished using energy, magnitude and spectrum. MFCC features are extracted from speech words. K- Nearest Neighbor (KNN) classifier is used as classifier. The developed system is tested for withdrawal slip filling application. The proposed system is designed and simulated using a MATLAB. The developed system achieved the recognition accuracy of about 91.5%. This system acts as a basis for real time speech recognition products, where input will never be an isolated word. 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.

REFERENCES

[1] TorbjBm Svendsent, Frank K Soong," On the Automatic Segmentation of Speech Signals", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '87,Vol. 12, pp. 77-80, 1987.

[2] M.Kalamani, Dr.S.Valarmathy, S.Anitha, R.Mohan "Review of Speech Segmentation Algorithms for Speech Recognition", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Vol. 3, Issue 11, pp.: 1572-1574, 2014.

[3] Y. Yang and X. Liu, " Re-Examination of Text Categorization Methods," Proc. SIGIR "99, pp. 42-49, 1999.

[4] Stefan Windmann, Reinhold Haeb-Umbach, "Approaches to Iterative Speech Feature Enhancement and Recognition", IEEE Transactions On Audio, Speech, And Language Processing, vol. 17, Issue- 5. pp. 974-984, 2009.

[5] Ang F , Sapporo, Tsutsui H , Miyanaga Y, "Time-varying LP cepstral features for improved isolated word speech recognition", IEEE International Conference on Digital Signal Processing (DSP), pp. 302 – 306, 2015.

[6] Principi E, Squartini S, Piazza F, "Power Normalized Cepstral Coefficients based super vectors and i-vectors for small vocabulary speech recognition", International Joint Conference on Neural Networks (IJCNN), pp. 3562 – 3568, 2014.

[7] Ren Wenxia, Zhang Huili, Lv Wenzhe, "Realization of Isolated-words Speech Recognition System", Pacific-Asia Conference on Circuits, Communications and Systems PACCS '09. pp. 353 – 355, 2009.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

[8] Lucey S, Sridharan S, Chandran V, "Improving visual noise insensitivity in small vocabulary audio visual speech recognition applications", Sixth International Symposium on Signal Processing and its Applications, Vol. 2 , pp. 434 – 437,2001.

[9] Chengalvarayan R, "Adaptation of quadratic trajectory segment models for small vocabulary speech recognition", Proceedings of International Conference on Information, Communications and Signal Processing ICICS, Vol. 2 pp.1007 – 1010, 1997.

[10] Kamm C A, Yang K M, Shamieh C R, Singhal S, "Speech recognition issues for directory assistance applications", Second IEEE Workshop on Interactive Voice Technology for Telecommunications Applications, pp.15 – 19,1994.

[11] Velez E, Cossette L, Cuperman V, "Development of a VQ-HMM continuous speech speaker-independent recognition system for small vocabularies", IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 1991, vol.2, pp. 469 - 472

[12] Walker K, deSilva C J S, Alder M, Attikiouzel Y, "A phonetically based small vocabulary automatic speech recognition system", Fourth IEEE Region 10 International Conference TENCON '89. Pp.:765 – 768, 1989.

[13] Brognaux S, Drugman T, "HMM-Based Speech Segmentation: Improvements of Fully Automatic Approaches", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 24, Issue 1, pp. 5 – 15, 2016.

[14] Receveur S, Weiss R, Fingscheidt T, "Turbo Automatic Speech Recognition", I EEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 24, Issue 5 , pp. 846 - 862, 2016.

[15] Anusuya MA, Katti SK "Mel frequency discrete wavelet coefficients for Kannada speech recognition using PCA" In Proceedings of international conference on advances in computer science ACEEE. pp. 225 – 227, 2010.

[16] Hemakumar G and Punitha P , "Speaker Independent Isolated Kannada Word Recognizer", published by P. P. Swamy and D. S. Guru (eds.), Multimedia Processing, Communication and Computing Applications, Lecture Notes in Electrical Engineering 213, Springer India, pp. 333-345, 2013.

[17] Renjith, S. ; Joseph, A. ; Babu, K.K.A.," Isolated digit recognition for Malayalam- An application perspective", International Conference on Control Communication and Computing (ICCC) , pp. 190– 193, 2013.

[18] Raji Sukumar, A. ; Sarin Sukumar, A. ; Firoz Shah, A. ; Anto P, B." Key-Word Based Query Recognition in a Speech Corpus by Using Artificial Neural Networks" Second International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN), pp.33 – 36, 2010.

[19] P. Punitha ; G., Hemakumar," Speaker Dependent Continuous Kannada Speech Recognition Using HMM", International Conference On Intelligent Computing Applications, pp. 402 – 405, 2014.

[20] Muralikrishna, H. ; Ananthakrishna, T. ; Shama, K." HMM based isolated Kannada digit recognition system using MFCC", International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp.730 – 733, 2013.

[21] Harisha S B, Amarappa S and Dr. S V Sathyanarayana "Spoken Digit Recognition Based on Support Vector Machine for Kannada Language", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X.Vol. 5, Issue 11, pp. 774-780, 2015.

[22] Harisha S B, Amarappa S and Dr. S V Sathyanarayana " 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,Vol. 4 Number 1, pp. 91-105, 2015.

[23] Hemakumar G and Punitha P, "Automatic Segmentation of Kannada speech signal into syllables and sub-words: Noised and Noiseless signals" International Journal of Scientific & Engineering Research, Vol. 5, Issue 1, pp. 1707-1711, 2014.

[24] 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, 2009.

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

[26] T. M. Cover, P. E. Hart, "Nearest Neighbor Pattern Classification", IEEE Trans. Inform. Theory, Vol. 13, Issue 1, pp. 21-27, 1967.