International Journal of Electrical and Computer Engineering (IJECE)
Vol. 9, No. 6, December 2019, pp. 4684~4695
ISSN: 2088-8708, DOI: 10.11591/ijece.v9i6.pp4684-4695

Continuous kannada speech segmentation and speech recognition based on threshold using MFCC and VQ

Vanajakshi Puttaswamy Gowda, Mathivanan Murugavelu, SenthilKumaran Thangamuthu
ACS College of Engineering, Visveswaraya Technological University, India

Article Info

Article history:

Received Dec 19, 2018
Revised Jun 20, 2019
Accepted Jun 30, 2019

Keywords:

Feature extraction
Minimum distance
Short term energy
Spectral centroid
Speech recognition
Speech segmentation
Zero crossing rate

ABSTRACT

Continuous speech segmentation and its recognition is playing important role in natural language processing. Continuous context based Kannada speech segmentation depends on context, grammer and semantics rules present in the kannada language. The significant feature extraction of kannada speech signal for recognition system is quite exciting for researchers. In this paper proposed method is divided into two parts. First part of the method is continuous kannada speech signal segmentation with respect to the context based is carried out by computing average short term energy and its spectral centroid coefficients of the speech signal present in the specified window. The segmented outputs are completely meaningful segmentation for different scenarios with less segmentation error. The second part of the method is speech recognition by extracting less number Mel frequency cepstral coefficients with less number of codebooks using vector quantization. In this recognition is completely based on threshold value.This threshold setting is a challenging task however the simple method is used to achieve better recognition rate.The experimental results shows more efficient and effective segmentation with high recognition rate for any continuous context based kannada speech signal with different accents for male and female than the existing methods and also used minimal feature dimensions for training data.

Corresponding Author:

P.Vanajakshi,
Research Scholar,
ACS College of Engineering,
Visveswaraya Technological University, India.
Email: vanaja_gowda@rediffmail.com

1. INTRODUCTION

Speech segmentation with recognition is challenging task now a days. Proposed speech segmentation in this paper is the processes of isolating the speech signal with context based information of the particular scenario. The segmentation of the speech based on context is to identify end points of the context information based on words, syllables or phonemes in natural langages.The speech segmentation is a subpart of speech recognition system. In natuaral language processing context, semantics, and grammer are very important to recognize the speech. The applications of automatic speech recognition system are used in broadcast news transcription, information extraction and retrieval, identity the speakers voice and allow the authenticated user to utilize the services and many more.

Higher lag is method is used to extract the features of speech signal with linear prediction [1]. The method gives two prediction errors, one is the ordinary convention linear prediction and other one is the delayed version with k number of samples of linear prediction. Further Combined Higher Lag Linear Prediction (CHLLP) model simultaneously by zero lag and higher lag prediction. In CHLLP model the cost function CHLLP model is equal to the cost function of conventional linear prediction if signal is completely periodic with fundamental period length equal to P number of samples and if m number of samples selected

from P.In CHLLP if zero lag prediction means m = 0, simultaneously weighted by scalar. So this method is a new spectral modeling method can be used as a modern speaker recognition system with different noisy environments.

Speaker independent for Telugu language for continuous words MFCC (mel frequency cepstral coefficients) [2] and DWT are used to extract the features of continuous speech then classified these extracted features using HMM based neural network. The MFCC and DWT combined methods are useful to extract features of the speech signal. Wavelet packet decomposition are used to remove the noise present in the feature of the speech signal.DWT isolate the higher and low frequency bands present in the speech signal. So high frequency bands are considered as a useful features. This process gives low word error rate in speech recognition system.

The method enhance the whisper recognition [3] by extracting a new robust cepstral features and preprocessing based on demising autoencoder. Teaser energy based cepstral features are more robust than MFCC for whispered description DDAER PECC feature extraction significantly improves the recognition rate compare MFCC, GMM, HMM. This proposed method improves more than 31% than traditional methods. Typically auto encoder has input layer which is the original feature vector one or more hidden layer which are Transformed features out of those hidden layer which matches input layer for reconstruction. TECC based features predicts the facts that Teoperator and gammet one filter bank to describe whishper characterstics. So, because of these the achieved word recognition rate is 93%.

Word boundary detection is used to separate the word from Gujarat Speech. This paper achieves end point detection [4] in Gujarath speech recognition system with the presence of background noise. It separates the silent portions of the speech.So that noise is reduced. This word boundary detection uses two algorithm to detect and point explicitly and implicitly. Explicit end point detection usedative before recognition and implicit end points are used after speech process to detect end point.this method able to detect weak fricative in signal to noise ratio condition.

The Hybrid is model used for maximum Gaussian mixture continuous Tamil Speech recognition [5]. This method improves accuracy upto 3% error rate up to 4% compare to the existing system. This model is used in speech to text conversion in various application. In this LPC, MFCC, LP are used to extract the features. This method is an unsupervised method to analysis of data and construction modeling. So this portion data points between zero and one. These values are assigned based on the clusters, centre and data points

To recognize isolated Kannada words trained HMM model and viterbi algorithm for decoding process [6]. MFCC are computed in frontend processing. This proposes to compare the performance of phone level and syllable level acoustic model for small to medium sized kannda language vocabulary. Average word recognition accuracy 97% for syllable level modeling, 98.6% for phone level meodeling. Speech coding setup has been done using HTK tool. The entire database training and testing samples are used to build by using HMM. Cofusion matrix is used to analyse and interpret the results at the word level.

HMM and Normal Fit [7] method is used for continuous speech recognition. Voice detection based on computing dynamic threshold and cepstrum coefficients are extracted as a feature of voice. The Baum-Walsh algorithm is used for trained database and Normal Fit tehnique is used to label the speech. This method tested for five languages. In an average accuration rate 95%. The experimental results shows that size of memory reduces because of Normal Fit values.

The MFCC is used [8] for feature extraction for training data. The Vector Quantization is used for clustering Speaker Independent Kannada Speech Recognition. VQ_1 and VQ_2 is used for clustering purpose. The Speech Recognition error decreases from 2.5 to 1.5. In case of VQ_1 and VQ_2 are the two clustering techniques, VQ_1 based on binary splitting algorithm and VQ_2 based on largest average distortion. Appling Linear discriminant analysis (LDA) [9] and maximum likelihood transformation on MFCC to extract features of speech and input these features to Convolution Neural Network (CNN) to improve robustness of speech recognition. This improves the recognition accuracy.

The proposed method organized into two parts; 1) Speech segmentation 2) Recognition; The first part of the paper is continuous kannada speech segmentation based on context and isolate kannada letters from continuous kannada speech which contains only kannada letters speech signal. This can be achieved by detection of voiced and unvoiced speech signal based on computing the average energy and spectral centroid of each frame of the Kannada speech signal. Average energy and spectral centroid coefficients are futher subjected to a median filter. The output of the median filter coefficient are used to set the thersholds. These thresholds are used to segment the continuous Kannada speech signal based on context. The second part of the paper is to determine the feature extraction of the segmented speech signal using threshold based MFCC and VQ in an automatic speech recognition (ASR) system. The threshold based MFCC and VQ is used to train speech data set. The methods uses less number of MFCC and less number of codebook of VQ gives better results than the existing methods.

2. PROPOSED METHODOLOGY

The first part of the proposed method contains continuous context based Kannada speech segmentation and isolated Kannada Akshara (means letter) from continuous Kannada speech which contains the utterance of Kannada Akshara only. Second part of method describes the continuous context based Kannada speech segments and isolated Kannada akshara speech recognition system using threshold based MFCC and VQ methods.

Speech segmentation part is this paper is carried out by detecting the presence of the voiced and unvoiced speech using average short time energy and spectral centroids. The median filter is used to smoothen the average short time energy and spectral centroids coefficients of the speech signal and threshold has been set based on the probability density function (pdf) of the output of the median filter coefficients.Then contex based speech segmentation performed based on the threshold levels.

2.1. Average short time energy (STE)

The Kannada speech signal is decomposed into a number of frames by multipling window function of length L using (1). Then each frame average short time energy [10-11] is computed using (2).

$$s_w(i) = s(k) \times w(i-k) \tag{1}$$

where $s(k)$ is kannada speech signal, $w(i)$ represents hamming window function, which is shifted across the speech signal to obtain frames and $s_w(i)$ is the windowed speech signal

$$E = \frac{1}{L}\sum_{i=1}^{L} s_w^2(i) \tag{2}$$

2.2. Spectral centroid features

The spectral centroid (SC) measures frequency and magnitude of the particular spectral bin using the Discrete Fourier Transform. The spectral centroid contains more energy above and below the fundemantal frequency, which is almost the average energy of the spectral bin. Usually the speech signal has asymmetric in naure about the pitch range. The accuracy of perception in speech signal in the form of ramp function, so that it gives more accurate perception in both lower and higher frequencies of the spectral bin. The each frame of the spectral centroid of size N is defined in (3)

$$SC = \frac{\left(\dfrac{\sum_{j=1}^{N} j \times M \times S_k(m)}{\sum_{j=1}^{N} S_k(m)}\right)}{\left(\dfrac{f_s}{2}\right)} \tag{3}$$

Where $S_k(m)$ is the FFT of windowed sequence of the speech signal of size N samples, $M = \dfrac{f_s}{(2 \times N)}$ is the width of the each spectral bin and f_s is the sampling frequency of the speech signal. The multiplication factor j in (3) refers to the perception of speech signal as a ramp function.

2.3. Median filter and thershold setting

The median filter is used futher to smoothen and retain any abrupt changes within $\frac{L}{2}$ of average energy and spectral centroid coefficients. Where L is the length of filter. In this paper length of the filter is 5. Since it is a non liear filter it will not smoothens the noise components presents in the average energy and spectral centroid coefficients. The median filter outputs are used to set the thresholds based on the probability density function (pdf) of the filter output coefficients. These thersholds are used to identify the context of the speech signal in appropriate manner. Energy threshold (ET) and spectral centroid threshold (ST) setting is required to segment the continuous kannada speech signal.Both the threshold can be computed by taking the histogram of the STE and Spectral Centroid of each frame. Two flags f1 and f2 are setting by comparing energy with ET and centroid with ST. Depending on the final flag, the speech segmentation is achieved based on the context of the scenario. Finally each frame of the speech is separated with voiced and unvoiced speech based on context.

2.4. Zero crossing rate and end point detection

Zero Crossing Rate gives information of rapidly changing of the speech signal from positive to negative. If more number zero crossings means the speech signal contains the high frequency information [12]. If it is less the signal contain low frequency information. Thus zero crossings is used in this paper to identify the voiced and unvoiced speech signal which is helpful to segments the given signal. In a given frame the speech signal is considered as non-stationary signal and It is defined in (4).

$$Z_{CR}(n) = \frac{1}{2N}\sum_{l=1}^{N} s(l) \times w(n-l) \quad (4)$$

Zero Crossing Rates (ZCR) is used to detect the voice activity in the speech signal, the signal whether it is a speech has spoken voice or silent. The ZCR used in this paper, to detect the end point of the speech signal within the context. Zero crossing rate is isolating the letter exactly from continuous speech. Zero crossing is playing important role in this aspect to separate individual letters. By masking unvoiced speech is considered as zero and voiced speech is maintained as it is in the original speech signal. Further each letters are isolated with their endpoints using short time energy and zero crossing rates.

3. SPEECH RECOGNITION

Context based recognition and Kannada Varnamala and Kannda alphabet recognitions are proposed from continuous Kannada sp each signal. Mel Frequency Cepstral Coefficients (MFCC) an Vector Quantisation (VQ) based feature extractions are proposed.

3.1. Mel frequency cepstral coefficients

Mel Frequency Cepstral Coefficients (MFCC) is one of the efficient and effective significant feature extraction method [13-15] used in speech recognition system. The Mel Frequency scale is non linear which represents based on the speech frequency range.Usually non linear frequency range eperception of speech signals are represented by the MFCC coefficients. The speech signal is passing through a band pass filter to obtain MFCC coefficients in which higher band frequencies and critical bands are enhanced and then pass through a inverse Fast Fourier Transform (FFT). So that the speech perception analysis are accuratelyconsider as a feature extraction for recognition system. The continuous speech signal is divided into N number of frames with m number of samples. This processing is done by windowing each individual frame with Hamming window technique. The signal $s(i)$ multiplying the window function $w(i)$ to obtain the windowed speech signal $s_w(i)$ as given in (5).

$$s_w(i) = s(i) \times w(i), \qquad 1 \le i \le m \quad (5)$$

The frequency analysis of windowed sequence is computed using discrete Fourier transform (DFT) in (6).

$$S_w(k) = \sum_{j=1}^{N} s_w(j) e^{\frac{-j2\pi jk}{N}} \quad (6)$$

The triangular band of frequencies are obtained using Mel-filter banks in (7).

$$f_{mel} = 2595 \log\left(1 + \frac{f}{700}\right) \quad (7)$$

3.2. Vector quantization (VQ)

Vector Quantization is one of the most important method of distance measure between the test data and trained data set in automatic speech recognition. Based on the minimum distance measurement, it is easy to recognise the test data present in the trained data set. VQ is the one of the method to reduce the number of significant dimensions of input data. So that, it matches the unknown models in a very simple manner by reducing the data. This VQ algorithm creates 8 number of dimensions in this paper, which produces a set of cluster centers spread the distance space depending on the speech.

Signal features.Then categarise any feature vector to one of these clusters and by using these cluster number as an input feature vector.Comparing [16, 17] two sequences of integers vectors than the entire original vectors.one of the additional advantage is to compute the distance between the pairs of clusters as

the Euclidean distance measure between their corresponding centers of the vectors. This is very simple to view the looked up table to measure the distance bwtween the clusters and not required any additional computations to measure distance.

In order to make the VQ method simple, a set of cluster centers is defined as a codebook because it produces feature vectors into single value. The codebook size refers to the number of clusters in the codebook. If any sort of information is lost when VQ is [18-20] method is used to encode an input vector sequence. Then grouping dissimilar points and representing every cluster member by computing the mean clusters. The loss of information is the difference between the original input vector sequence and the quantised vector sequence. The mean value of this difference is the VQ distortion. By increasesing the size of the codebook leads to decreases in VQ distortion.

3.3. Threshold setting

The output of the codebook is used to set the threshold value to recognize wether test speech signal is present or not in the training dataset. The minimum value in the code book is used as threshold value but this threshold should be less the half of the average value in the codebook, then only the test speech signal is allowed to test in the training data set otherwise test speech signal is not present in the training data set. Once the test speech signal is allowed to test, it looks only the minimum distance vector, the minimum distance vector speech is recognize as a test signal speech.

3.4. Proposed model algorithm

Context based voice detection:

Step 1: Input data continuous kannada speech signal

1. The speech signal of sampling frequency f_s Hz and hamming
2. window length (N) = step size = 0.050 $\times$ f_s .
3. Compute number of frames of speech using (1) by shifting the window the across the entire speech signal.

Step 2: Compute the average energy of each frame using equation (2).

Step 3: Compute $2\times N$ point FFT of windowed sequence of each frame.

1. Consider only N point FFT coefficients to reduce higher spectral components.
2. Spectral centroid 'C' of each windowed sequence using equation (3)

Then finally centroid C is $C = \frac{C}{\left(\frac{f_s}{2}\right)}$.

Step 4: Filtering Filter the average energy sequence and centroid sequence using median twice of filter length of five and compute

E_{mean} = mean($E_{filtered}$), C_{mean} = mean $(C_{filtered})$

1. Find the threshold using pdf of enegy and centroid sequence.
2. Compute the threshold as the weighted average between two first pdf local maxima then threshold energy= $\frac{E_{mean}}{2}$.
3. Similarly step3 is repeated for centroid sequence.

Step 5: Set the Threshold values

1. Set flags f_1 and f_2 .
2. $f_1 = E \leq$ threshold energy.
3. $f_2 = C \leq$ .threshold Centroid.
4. $f = f_1 \& f_2$.

Step 6: Speech detection.

1. Initialise count=1, flag=1.
2. Set start limit.
3. Increase overall counter. Increase counter of the current speech segment.
4. If atleast one segment has been found in the current loop set end counter then increase overall counter.
5. Merge overlapping segments.
6. Plot the segmented speech signal by representing in red colour and play each segment. Finally written each segment as .wav file.

7. Segment the continuous kannada speech signal based on context using flags f_1 and f_2.
8. Isolate the kannada letter speech signal using ZCR from continuous kannada speech based on context which contains only kannada letter speech signal.

Plot the isolated kannadaletter speech signal by representing in red colour and play isolated kannada speech.Finally written each letter speech as .wav file.

Speech Recognition

Step 1: Take segmented speech signal as a training dataset
Step 2: Apply the Fourier transform segmented speech signal
Step 3: Map the log amplitudes of the spectrum obtained above onto the Mel scale, using triangular overlapping windows.
Step 4: Take the Discrete Cosine Transform of the list of Mel log-amplitudes, as if it were a signal.
Step 5: The MFCCs are the amplitudes of the resulting spectrum.
Step 6: Calculate MFCC Coefficient for training dataset with frequency rate 10
Step 7: Generate code book for each segmented MFCC coefficients using vector Quantization (VQ) with 8 number using equidistance and keep these codebooks as a training data set.
Step 8: Repeat step6 and step7 for test speech signal
Step 9: Set the threshold by computing the the minimum value of codebooks in training data set.
Step 10: Compute the average value of codebooks in training data set.
Step 11: If threshold value is less than half of the average value**,** then it check the test signal in the training data set otherwise test speech signsl is not recognized.
Step 12: Compute distance between test data with training data.
Step 13: The minimum distance vector speech in the training data set is considered as recognized
Step 14: Stop

4. RESULTS AND DISCUSSIONS

4.1. Segmentation

The Figure 1 shows that continuous original kannada Speech signal segmented into four parts. The short time energy of speech signal and corresponding spectral centroid of each segmented output mentioned with green colour and its filtered output with blue colour. The median filter output is completly smoothen so that any distortion present in the speech signal completely eliminated. The segmented speech signal completely isolated with voiced and unvoiced with respect to the particular scenario of that context. The Figure 2, shows the corresponding kannada Speech signal text. Each segmented output is completly meaningful with respect to the kannada syntatic, semantic and grametic rules which is mentioned as in Figure 3 of (a-d) using unicode of kannada language. The Figure 4 illustrates some of the words present in the segmented speech signal text written using Unicode of kannada language with matlab R2014a.

The Table 1 gives the context based continuous kannada speech signal segmentation with different accent of different signal size of male and female speech. The segmentation algorithm tested nearly 100 different kannada speech signal of female and male with different accent. This algorithm gives 0.01 % of segmentation error rate. Only three different speech of different duration is listed in Table 1. The original segmentation with contex based is almost nearly equal to practical segmentation of correct segments. The number of missed and extra segmentation almost nil and error rate is almost neligible. So, the algorithm which has been mentioned in this paper gives correct segmentation with less error rate. The Table 2 shows that segmentation of context based with different vocabulary for male and female of different accent gives depending on the vocabulary size the segmentation accuracy decreases as vocabulary size increases but proposed algorithm gives better segmentation accuracy even for large vocabulary size.

The Figure 5 shows the continuous kannada letters speech signal and segmentation of continuous kannada letters speech signal shows the continuous kannada letters speech signal which contains all 52 letters of kannada language. This speech signal letters are isolated in Figure 5 (a-h) which are mentioned only seven letters with red mark in speech signal.

The Table 3 contains the segmentation of the kannada akshra from continuous kannada akshra speech signal of male and female. Segmentation of kannada akshra in this case is the isolation of kannada akshra. The continuous speech signal contains total 52 kannada akshar's. the original segments must be 52 akshar's but isolated kannada letters from the algorithm which has been mentioned gives more than 52 akshar's which contains correct 52 akshar's segments and extra segments which are not required segments. The error occurs due to the extra segments which is redundant, but the missing segments nil for both male and female speech signal.

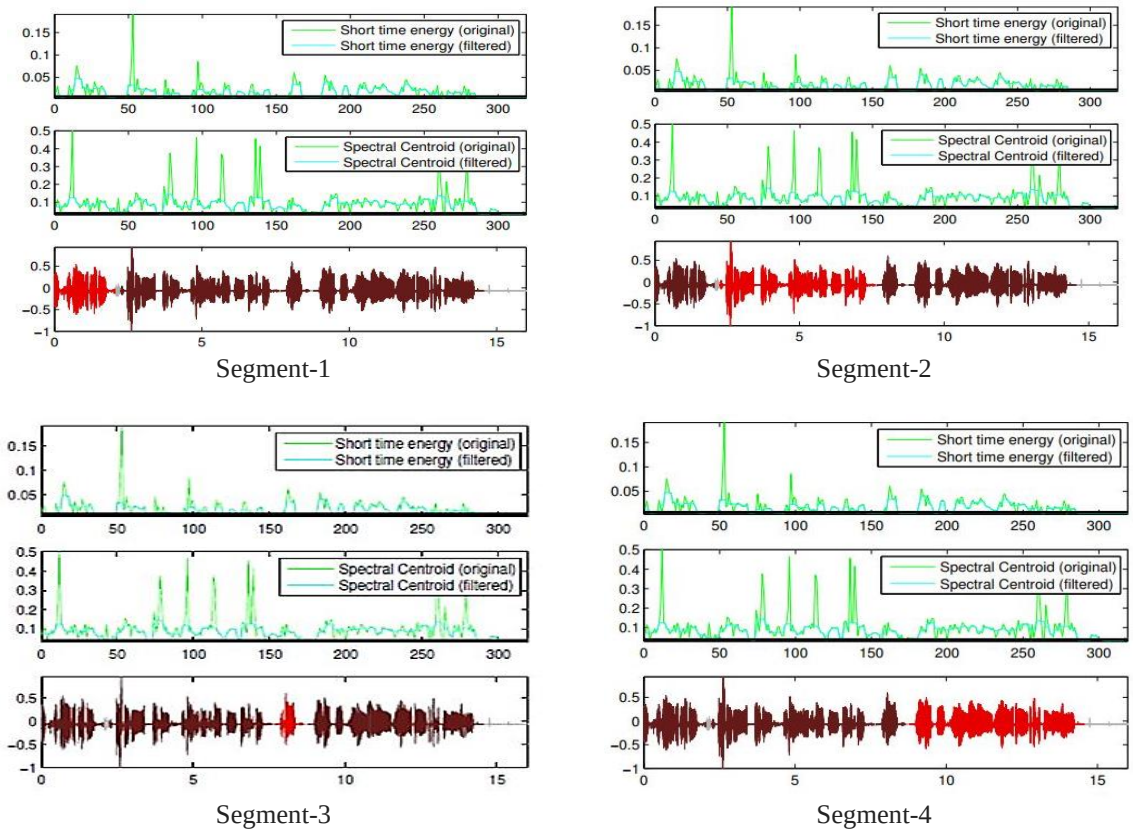

Figure 1. Continuous Kannada speech Signal segmented into four parts

ಆ ಸರೋವರದ, ಪರ್ವತರಗಳನ್ನು ತಿಳಿಸಿದನು, ಆ ಸರೋವರದ ಹೆಸರು ಪಂಚಸ್ವರ,
(aa sarovarada parvatharagalannu thilisidanu ,aa sarovarada hesaru panchaswara,)

ಅದನ್ನು ಮಂಡಾಕಿನಿ ಎಂಬ ಮಹಾಮುನಿಯು ತನ್ನ ತಪಶಕ್ತಿಯಿಂದ ನಿರ್ಮಿಸಿದನು.
(adannu mandarkini yamba mahamuniyu thanna thapashakthiyinda nirmisidanu)

Figure 2. Original Kannada Speech signal text

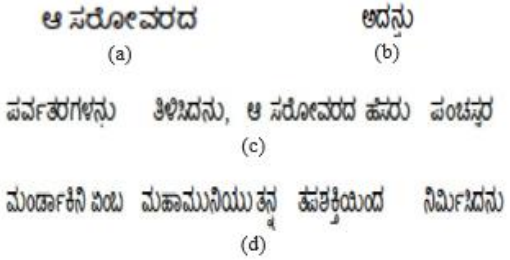

Figure 3. Kannada Speech signal is segmented into four parts (a), (b), (c) and (d)

Kannada word	Unicode
ಸರೋವರದ (Sarovarada)	0CB8,0CB0, 0CCB, 0CB5, 0CB0, 0CA6
ಪರ್ವತ (Parvatha)	0CAA,0CB0, 0CCD, 0CB5, 0CA4
ಪಂಚಸ್ವರ (Panchaswara)	0CAA, 0C82, 0C9A, 0CB8, 0CCD, 0CB5, 0CB0
ಅದನ್ನು (Adannu)	0C85, 0CA6, 0CA8 , 0CCD, 0CA8, 0CC1
ಮಂಡಾಕಿನಿ (Mandarkini)	0CAE , 0C82, 0CB0 ,0CCD, 0CA1 ,0CBE,0C95, 0CBF, 0CA8, 0CBF
ಶಕ್ತಿ (Sakthi)	0CB6 , 0C95, 0CCD, 0CA4, 0CBF

Figure 4. Unicode representation of Kannada text

Table 1. Context based continuous Kannada Speech segmentation with different accent

Speakers	No. of different speakers	No. of original segments	No. of correct segments	No. of missed segments	No. of extra segments	Error Rate
Female with	S1	4	4	Nil	Nil	Nil
signal_1	S2	4	4	Nil	Nil	Nil
	S3	4	4	Nil	Nil	Nil
Male with	S1	4	4	Nil	Nil	Nil
signal_1	S2	4	4	Nil	Nil	Nil
	S3	4	4	Nil	Nil	Nil
Female with	S1	10	10	Nil	Nil	Nil
signal_2	S2	11	12	Nil	1	1
	S3	10	11	Nil	1	1
Male with	S1	11	12	Nil	1	1
signal_2	S2	10	11	Nil	1	1
	S3	11	10	1	Nil	1
Female with	S1	62	62	Nil	Nil	Nil
signal_3	S2	63	62	Nil	1	1
	S3	63	62	Nil	1	1
Male with	S1	62	62	Nil	Nil	Nil
signal_3	S2	63	62	Nil	1	1
	S3	63	62	Nil	1	1

Table 2. Segmentation of context based Kannada speech signal with different vocabulary

Speakers	Number Speakers	Size of vocabulary(segments)	Segment accuracy (%)	Error rate (%)
Female	50	100	99	1
Male	50	100	99	1
Female	50	500	97	3
Male	50	500	98	2
Female	50	1000	96	4
Male	50	1000	97	3

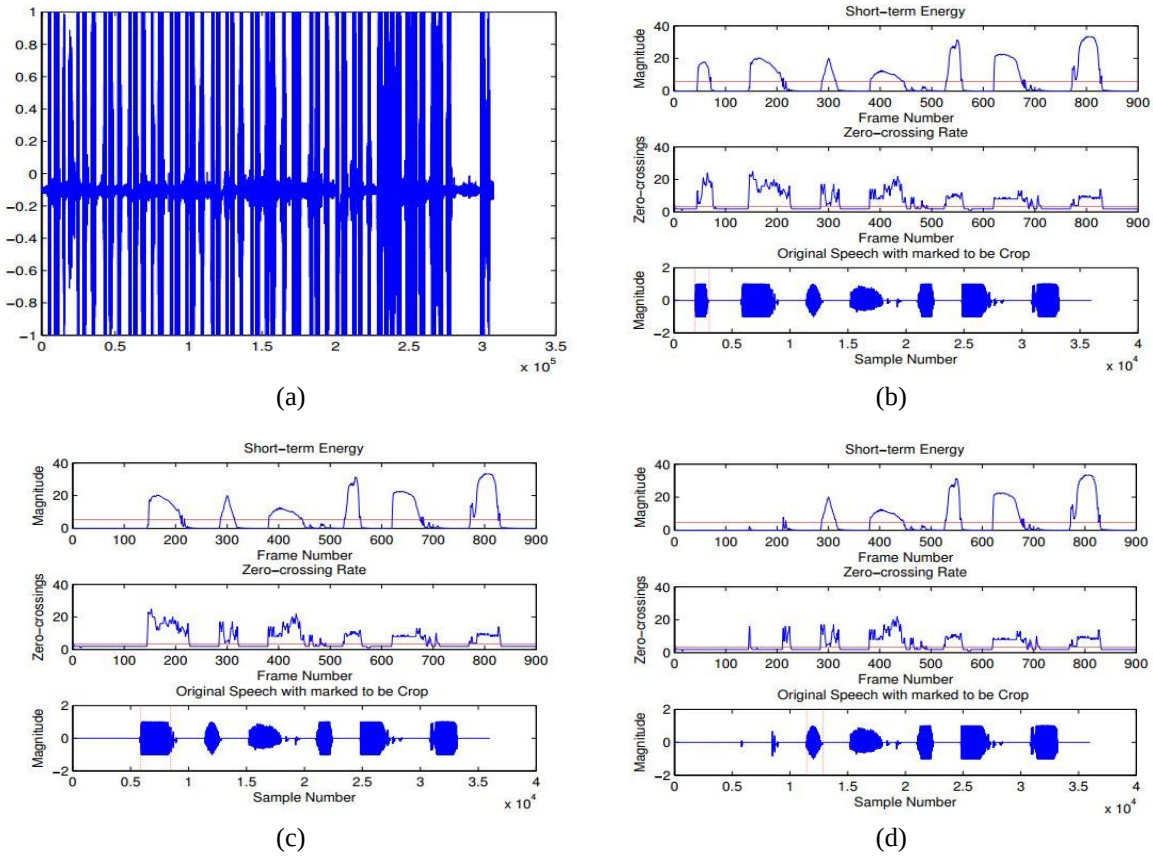

Figure 5. (a) Continuous kannada letters speech signal, (b) Akshara 'a', (c) Akshara 'aa', (d) Akshara 'e'

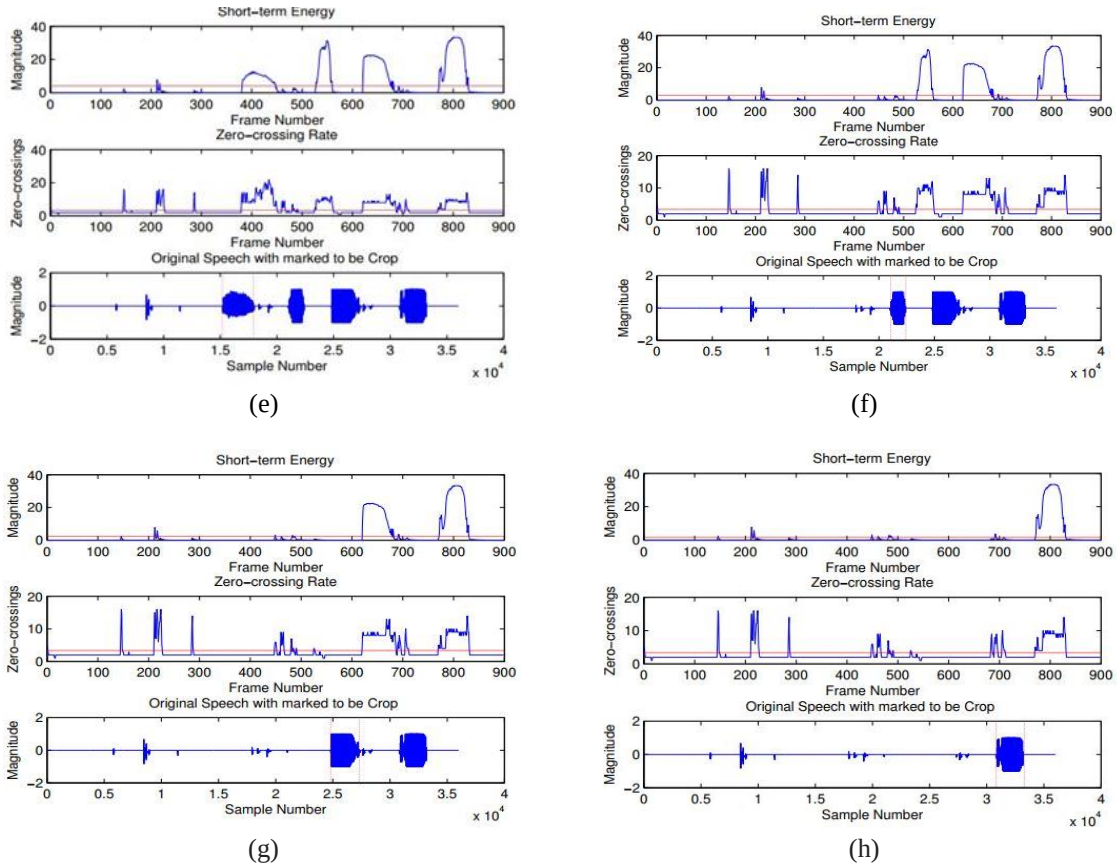

Figure 5. (e) Akshara 'ee', (f)Akshara 'u', (g), Akshara 'uu', (h) Akshara 'ru'

Table 3. Kannada Akshara (letters) Speech Segmentation with different accent

Speakers	No.of different Speakers	No.of Original segments	No.of Correct segments	No.of missed segments	No.of extra segments	Segmentation Error
Female	1	52	57	Nil	5	5
	2	52	54	Nil	2	2
	3	52	56	Nil	4	4
Male	1	52	55	Nil	3	3
	2	52	57	Nil	5	5
	3	52	56	Nil	4	4

4.2. Recognition

The Table 4 illustrates the segmented Kannada letters speech recognition for male and female. The isolated Kannada akshara speech signal is used as a traing data set for the recognition system. The recognition system gives '1' for the the recognised segment which is present in the trained data set i.e the test segment otherwise it is '0'.

Table 4. Isolated Kannada (Akshara) letters speech recognition

Speakers	Test Segments	Training Segment							
		ಅ	ಆ	ಇ	ಈ	ಉ	ಊ	ಎ	ಋ
Female	ಅ	1	0	0	0	0	0	1	0
	ಆ	0	1	0	0	0	0	0	0
	ಋ	0	0	0	0	0	0	0	1
	ಎ	0	0	0	0	0	0	0	0
Male	ಇ	0	0	1	0	0	0	0	0
	ಈ	0	0	0	1	0	0	0	0
	ಋ	0	0	0	0	0	0	0	1
	ಔ	0	0	0	0	0	0	0	0

Table 5 contains the context based continuous Kannada speech recognition. The segmented speech signal is used as a traing data set for the recognition system. The recognition system gives '1' for the recognized segment which is present in the trained data set i.e the test segment. Otherwise it is '0'. This recognition system gives better recognition rate.

The Table 6 shows the comparison of speech recognition of different feature extraction methods of different languages which are referred in these papers [21, 22]. This table mainly compares the MFCC along with other technique of feature extraction methods used for test and traing data set which are are used as a input to the recognition system. Speech signal of different languages depends on the syntax and symantic analysis of that language. Especially the Indian languages have large sets of vocabulary. The recognition of Indian language speech signal as shown in Table 5 with different recognition rate. The proposed method which has been used gives better recognition rate with less number of MFCC coefficients than the other language. The Figure 6 shows the graphical representation of recognition accuracy of different languages [23-25]. The method which is used in this paper for Kannada language speech recognition shows better accuracy rate than the existing methods.

Table 5. Context based continuous Kannada speech recognition

Speakers	Test Segments	Training Segments							
		Seg1	Seg2	Seg3	Seg4	Seg1	Seg2	Seg3	Seg4
Female	Seg1	1	0	0	0	1	0	0	0
	Seg2	0	1	0	0	0	1	0	0
	Seg3	0	0	1	0	0	0	1	0
	Seg4	0	0	0	1	0	0	0	1
Male	Seg1	1	0	0	0	1	0	0	0
	Seg2	0	1	0	0	0	1	0	0
	Seg3	0	0	1	0	0	0	1	0
	Seg4	0	0	0	1	0	0	0	1

Table 6. Recognition accuracy of different feature extraction methods of different languages

Languages	Feature Extraction Methods	Percentage of Recognition rate
Indonesian	MFCC, CMN_MFCC, PLP	88,90.88
Korean	MFCC (GMM)	78
English	MFCC	95
Arabic	MFCC	91
Indian (Tamil)	LPC+MFCC	89
Indian (Telugu)	LPC+MFCC	90
Indian (Malayalam)	LPC+MFCC	90
Indian (Kannada)	LPC+MFCC	91
Indian (Kannada)	Proposed	98

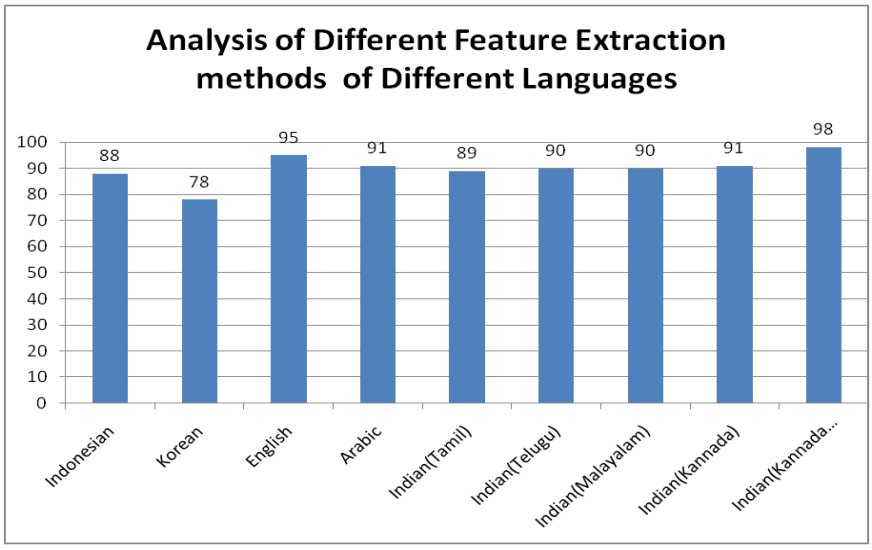

Figure 6. Recognition accuracy of different feature extraction methods of different languages

5. CONCLUSION

Segmentation and recognition of Kannada speech signal gives better results in this paper because of the proposed methods are efficient and effective based on the speech signal of different scenario. The methods contributes good segmentation with respect to the context of Kannada speech signal with segmentation errors are very less, but it depends on the vocabulary size. For large vabulary size this proposed method gives good segmentation with less missed segments. The speech recognition system is based on threshold with minimum number of MFCC and minimum number of codebook of VQ features gives better recognition rate forkannada speech signal with different scenario. The recognition system produces good recognition rate also works for different accents for male and female. In future there are more challenges, to reduce errors by using using different segmentation and recognition techniques.

REFERENCES

[1] Alku P and Saeidi R, "The Linear Predictive Modeling of Speech From Higher-Lag Autocorre-lation Coefficients Applied to Noise-Robust Speaker Recognition," *IEEE/ACM Transactions on Audio,Speech, and Language Pro- cessing*, Vol.25, No.8, pp.16061617, 2017.

[2] Kumar A.P, Roy R. Rawat S. Sudhakaran P., "Continuous Telugu Speech Recognition through Combined Feature Extraction by MFCC and DWPD Using HMM based DNN Techniques," *International Journal of Pure and Applied Mathematics*, Vol 114, No.11, pp.187–197, 2017.

[3] Grozdi T, Jovici I T, (2015) "Whispered speech recognition using deep denoising autoencoder and inverse filtering," *IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP), INPROCEEDINGS*, pp.2313–2322, 2017,

[4] Vijayendra A. D. and Thakar V. K. "Word boundary detection for Gujarati speech recognition using in-ear microphone," *2016 1st India International Conference on Information Processing (IICIP)*, @INPROCEEDINGS, pp.1–6, 2017.

[5] Kalamani M. Valamathy S. and Anitha S. "Hybrid Speech Segmentation Algorithm for Contin- uous Speech Recognition," *International Journal on Applications of Information and Communication Engineering (IJAICE)*, Vol.1, Issue. 1 , pp.2204–2210, 2015.

[6] Ananthakrishna "Study of sub-word acoustical models for Kannada isolated word recognition system," *International Journal of Speech Technology*, *Springer*, DOI 10.1007/s10772-016-9374-0, pp. 817– 826, 2016.

[7] Hemakumar G. and Punithavalli M. Thippeswamy K. "Large Vocabulary in Continuous Speech Recognition using HMM and Normal Fit," *International Journal of Computer Trends and technol- ogy (IJCTT*), Vol.42, No.2, pp.102–107, 2016.

[8] Anusuya M.A and Katti S.K "*Speaker Independent Kannada Speech Recognition using Vector Quantisation*," MPGI National Multi Conference 2012 (MPGINMC), @INPROCEEDING (IJCA), pp.33– 35, 2012.

[9] Hilman F. Pardede, Asri R. Yuliani, Rika Sustika, "Convolutional Neural Network and Feature Transformation for Distant Speech Recognition," *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 8, No. 6, December 2018, pp. 5381-5388 ISSN: 2088-8708

[10] S. Kumar S. Phadikar K. Majumder "*Modified segmentation algorithm based on short term energy and zero crossing rate for maithili speech signal*," Accessibility to Digital World (ICADW) 2016 International Conference on. IEEE pp. 169-172, 2016.

[11] T. T. Swee, S. H. S. Salleh and M. R. Jamaludin, "*Speech pitch detection using short-time energy*," International Conference on Computer and Communication Engineering (ICCCE'10), Kuala Lumpur, pp. 1-6, 2010.

[12] Pafan Doungpaisan , Anirach Mingkhwan, "Query by Example of Speaker Audio Signals using Power Spectrum and MFCCs," *International Journal of Electrical and Computer Engineering (IJECE*),Vol. 7, No. 6, pp. 3369 – 3384, December 2017.

[13] E. D. Dimaunahan, A. H. Ballado, F. R. G. Cruz and J. C. Dela Cruz, "*MFCC and VQ voice recognition based ATM security for the visually disabled*," 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM),pp.1-5, 2017.

[14] Lee S. J.and Kang B. O.and Chung H. and Park, J. G. (2015) "*A useful feature-engineering approach for a LVCSR system based on CD-DNN-HMM algorithm*," 23rd European Signal Processing Conference (EUSIPCO), @INPROCEEDINGS, pp.1421–1425, 2015.

[15] Rabiner L.Juang B.H. Yegnanarayan B. (2009) "Fundamentals of Speech Recognition," Pearson Pub- lications, 2009.

[16] Kalamani M. Valamathy S. and Krishnamoorthi, M. "Hybrid Modeling Algorithm for Con- tinuous Tamil Speech Recognition", *International Journal of Computer and Information Engineering (IJCAIE*), Vol.8, No.12, pp.2204–2210, 2014.

[17] Ali O. Abid Noor," Robust speaker verification in band-localized noise conditions," *Indonesian Journal of Electrical Engineering and Computer Scie*nce, Vol. 13, No. 2, pp. 499-506, February 2019.

[18] Satyanand Singh, "The role of speech technology in biometrics, forensics and man-machine interface", *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 9, No. 1, pp. 281-288, February 2019.

[19] Mijanur Rahman Md. Farukuzzaman Khan Md. and Amin Bhuiyan Md. "Continuous Bangla Speech Segmentation, Classification and Feature Extraction," *IJCSI International Journal of Computer Science*, Issues, Vol. 9, Issue 2, No.1, March 2012, ISSN (Online): 1694-0814, 2012.

[20] Satriawan, C. H. and Lestari, D. P. (2014) "Feature-based noise robust speech recognition on an In- donesian language automatic speech recognition system," *International Conference on Electrical Engineering and Computer Science (ICEECS)*, @INPROCEEDINGS, pp.42–46, 2014.
[21] Sangeetha J. Jothilakshmi S. Devendrakumar R.N.(2015) "Efficient Continuous Speech Recognition Approaches for Dravidian Languages", DOI 10.5013/*IJSSST*.a.15.02.03, 2015.
[22] Bishnu (2012) "Recognition of Isolated Words using Features based on LPC, MFCC, ZCR and STE, with Neural Network Classifiers," *International Journal of Modern Engineering Research (IJMER)*, Vol.2, Issue.3 pp 854–858, 2012.
[23] Sharma S.and Kumar M. and Das P. K.(2015) "A Technique for dimension reduction of MFCC spectral features for speech recognition", *International Conference on Industrial Instrumentation and Control (ICIC)*, @INPROCEEDINGS, pp.99–104, 2015.
[24] Abdo ,M. S and Kandil, A. H.and Fawzy ,S. A. "*MFC peak based segmentation for continuous Arabic audio signal*", 2nd Middle East Conference on Biomedical Engineering, @INPROCEEDINGS, pp.224–227, 2014.
[25] Londhe, N. D.and Kshirsagar,G. B. "*Continuous speech recognition system for Chhattisgarhi*," 2017 International Conference on Communication and Signal Processing (ICCSP), @INPROCEED- INGS, pp.0365–0369, 2017.

BIOGRAPHIES OF AUTHORS

Mrs. P. Vanajakshi has received her B.E degreein Computer Science and Engineering from Mysore University in 1997 and M.Tech degree in Computer Science and Engineering from Visveshwaraya Technological University, Belaghavi in 2007.She is currently pursuing her Ph.D in Visveshwaraya Technological University. Her area of interest includes Signal Processing, Speech Processing, Natural Language Processing, Artificial Intelligence. She is working since from 2000 at Vivekananda Institute of Technology, Bengaluru and she has 19 years of teaching experience. She is the life time member of ISTE, MIE and IAENG professional societies.

Dr. M. Mathivanan born in Tamilnadu, India and obtained his B.E degree in ECE from University of Madras, M.E degree in Applied Electronics and Ph.D in ECE from Anna University, Chennai. He has 18 years of teaching experience and more than 10 years of research experience. He has published 10 research articles in International Journals and more than 10 National & International Conferences. His area of interest includes Signal Processing, Speech Processing and Embedded Systems. Presently he is working as an Associate Professor, Department of ECE, A.C.S.College of Engineering, Bangalore affiliated to VTU, Belagavi. He has reviewed many research papers in various Journals and Conferences. He has acted as Indian Examiner for the research scholars of few Universities. He is the life time member of ISTE and IETE professional societies.

Dr T. Senthil Kumaran received his B.E. degree from University of Bharathiar University, Coimbatore 1998, M.E. degree from the same University, 2001. He received Best Project Award in his B.E Degree studies. His received Doctoral Degree from Anna University, 2014. He has got teaching, research and administrative experience of more than 18 years in various engineering colleges, autonomous institutions and universities. He has published more than 20 papers in national and international conferences and in international journals. He is working as scientific and editorial board member of many journals. He has reviewed dozens of papers in many journals. He received Outstanding Faculty in Computer Science and Engineering - 2017 by Venus International Faculty Awards. . He is a Life member of the ISTE, Senior Member IACSIT, Life Member IAENG, Member ICST, IAES