CURVELET BASED SPEECH RECOGNITION SYSTEM IN NOISY ENVIRONMENT: A STATISTICAL APPROACH

Nidamanuru Srinivasa Rao[1], Chinta Anuradha[2], Dr. SV Naga Sreenivasu[3]

[1]Scholar, Department of CSE, Acharya Nagarjuna University, Guntur, AP
[2]Asst.Prof, Department of CSE, VR Siddhartha Engineering College, Vijayawada, AP
[3]Professor, Dept. of CS, Narasaraopeta, Guntur, AP

ABSTRACT

Speech processing is considered as crucial and an intensive field of research in the growth of robust and efficient speech recognition system. But the accuracy for speech recognition still focuses for variation of context, speaker's variability, and environment conditions. In this paper, we stated curvelet based Feature Extraction (CFE) method for speech recognition in noisy environment and the input speech signal is decomposed into different frequency channels using the characteristics of curvelet transform for reduce the computational complication and the feature vector size successfully and they have better accuracy, varying window size because of which they are suitable for non –stationary signals. For better word classification and recognition, discrete hidden markov model can be used and as they consider time distribution of speech signals. The HMM classification method attained the maximum accuracy in term of identification rate for informal with 80.1%, scientific phrases with 86%, and control with 63.8 % detection rates. The objective of this study is to characterize the feature extraction methods and classification phage in speech recognition system. The various approaches available for developing speech recognition system are compared along with their merits and demerits. The statistical results shows that signal recognition accuracy will be increased by using discrete Curvelet transforms over conventional methods.

KEYWORDS

Speech Signals, Curvelets, HMM, Bayesian Networks, Recognition System.

1. INTRODUCTION

Automatic Speech Recognition (ASR) is the capability of machines to recognize words and phrase from spoken language and translate them into machine readable set-up [2]. The major interest of speech recognition is to build up methods and system for speech input signal to machine. Speech is the prime that means communication between humans of this medium motivates research efforts to permit speech to become a feasible human computer interaction [3]. Speech signals are non-stationary in nature; speech recognition is a complex job due to the multiple features like gender, pronunciation, emotional state, accent, articulation, nasality, pitch, and volume. Due to the background noise and other types of disturbances also makes a speech processing process is too complex. The throughput of a speech processing process is measured in terms of detection accuracy [4]. The fast evolution of computer hardware, software and speech recognition method is fairly turned as major expertise in processing of computer information. It is generally used in voice-activated telephone exchange, weather forecasting, banking services, computer aided language learning, video games, transcription, robotics, audio and video search, household applications and language learning applications etc [4].

DOI: 10.5121/ijcsit.2018.10304

Speech recognition system is separated into two segments, feature extraction and feature classification. Feature extraction method acts a crucial role in speech recognition process. Isolated word/sentence recognition requires extraction of characters from the recorded utterances followed by a training period [6]. The most extensively used feature extraction methods are Discrete Fourier Transform (DCT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Linear Predictive Coding (LPC), Mel Frequency Cepstral Coefficients (MFFCs), RASTA filtering and Hybrid Algorithm DWPD. A relatively new technique used in the field of signal processing for feature extraction is the curvelet technique. Classification methods used in speaking identification process, it include Gaussian Mixture Models (GMMs), Vector Quantization (VQ), HMMs and ANNs [6]. The current speech recognition scheme is based on the Hidden Markov Models for the spectral features of speech frame. The main reason why HMM is well-known and widely used is that HMM has parameters that can be automatically learned or trained and the techniques that are used for learning are easy and are computationally feasible to use. The following Figure 1 describes the primary task of speech identification system in effortless equations which consist of feature extraction, database, network training, testing and decoding. The detection process is shown below:

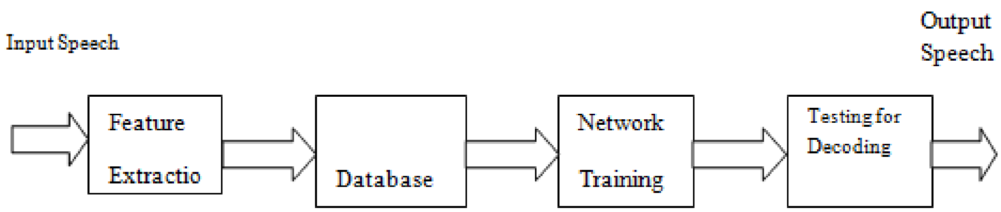

Figure 1. Basic model of speech recognition system

In this research, a novel method for speaking recognition is presented. This method is based on the extraction of the curvelet parameters from the original speech signal, curvelet coefficients are calculated for clean and noisy speech signals respectively. A Hidden Markov Model is used for classification and recognition. The objective of this method is to enhance the better performance of the new method for feature extraction over traditional approaches.

The rest of the paper is organized as follows. Section 2 discusses the Speech Recognition System. Curvelet Transform Technique for feature extraction is discussed in Section 3. Section 4 discusses the Hidden Markov Model for classification and pattern recognition. In Section 5, the proposed methodology is introduced and also gives the statistical results. Finally, Section 6 summarizes the concluding remarks.

2. SPEECH RECOGNITION SYSTEM

Speech recognition is the system for automatically detecting the spoken words of person based on data in speech signal. The progress of speech recognition system needs concentration in defining dissimilar types of speech classes, feature extraction methods, speech signal classifiers and performance evaluation. The major criterion for the good quality speech identification system is the selection of feature extraction method which acting a major role in accuracy of the system. Speech recognition process can be separated into dissimilar classes based on type of speech utterances are capable to recognize. Mathematically Speech signal system is defined as follow: Let the clean speech y (e) is degraded by additive noise u (e), so the corrupted signal x (e) can be expressed as

$$x(e) = u(e) + y(e) \quad \ldots\ldots (1)$$

The objective of speech restoration is to estimate x (e) from y (e) when there is little or no prior knowledge of u (e). The continuous speech waveform is first divided into frames. For each spoken word, let O = o1o2 · · · oτ be a sequence of parameters vectors, where ot is observed at time t ∈{1 . . . τ}. Given a dictionary D with words wi ∈D, the recognition problem is

$$W^{\sim} = \arg max_{W}^{P} (W|O) \quad (2)$$

Where W˜ is the recognised word, P is the probability measure and W = w1 ..... wk a word sequence and transform (1) in a suitable calculable form as:

$$P(W|O) = \frac{P(O|W)P(W)}{P(O)} \quad (3)$$

Where P(O|W) represents the acoustic model and P(W) the language model; P(O) can be unheeded. Thus, for a specified set of prior probabilities P(W), the probable spoken words based only on the probability P(O|wi). On Simplification (1) becomes

$$W^{\sim} = \arg max_{W}^{P} (O|W)P(W)^{s} |W|^{p} \quad (4)$$

In the log domain, the total likelihood is calculated as

$$\log_{|W|} W^{\sim} = \log_{|W|} P(O|W) + s \log_{|W|} P(W) + p \quad (5)$$

Where |W| is the length of the word sequence W, s is the language design scale factor (LMSF) and p the word insertion penalty (WIP).

Different classes of speech recognition system are i).Isolated Words: These systems require each utterance to contain lack of an audio signal on both sides of constructed sample window. It accepts sole word at a specified time. ii) Connected Words: It permit separate utterances to be spoken together with a minimum pause between them. iii).Continuous Speech: these systems allow people to speak almost in a natural way. iv). Continuous speech recognizers are hard to make as they need extra effort to determine word boundaries. v). Spontaneous Speech: Speech recognition process with spontaneous speech should be able to handle different natural speech character such as words run together with slight stutters.

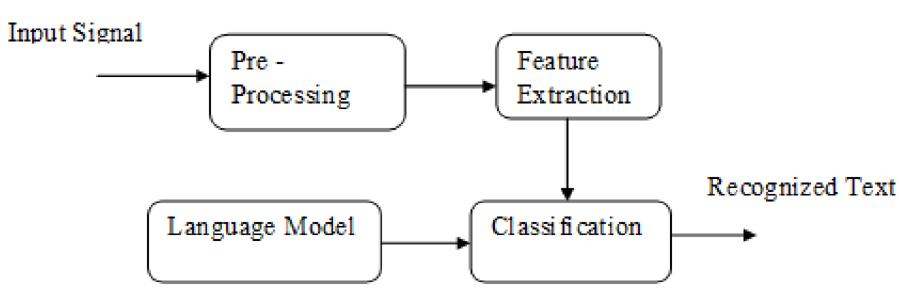

Figure 2. A Basic Model of Speech Recognition System

Figure 2 describes a fundamental model for speech recognition system; it represents different stages of a system consist of pre-processing, feature extraction and classification. The pre-processing stage indicated the input signal. After pre-processing, feature extraction stage extracts necessary vectors for modelling stage. These vectors necessity be robust for better accuracy.

Classification stage recognize the speech text using extracted features, where it contains syntax and semantics related to language accountable that helps classifier to recognize the input [1].

3. CURVELET BASED FEATURE EXTRACTION

Feature extraction is a central phase in speech computation and it is responsible for converting the speech signals into stream of feature vectors coefficients. These coefficients consist of information which is necessary for the identification of a given utterance. Every speech has different unique attributes contained in spoken words. These attributes can be extracted from a wide range of feature extraction methods and employed for speech recognition process [7]. The feature vector of speech signals are naturally extracted using spectral analysis methods such as Mel- frequency cepstral coefficients, linear predictive coding, curvelet and wavelet transforms etc. Using curvelets, the computational complexity, the feature vector size are reduced, gives better accuracy, varying window size because of which they are suitable for non –stationary signals. Thus curvelet transform is an elegant tool for the analysis of non-stationary signals like speech. Aamong all these techniques curvelet transform gives better system accuracy.

Discrete Curvelet transform (DCT) has good signal properties, it is applicable for many real signals and it is also computationally efficient. To obtain a time-scale representation of the signals DCT uses digital filtering. Let f (x1, x2), be function y1 = 0, 1,..., N1 - 1, x2 = 0, 1,..., N2 - 1, based on this function, we define the discrete Fourier transform [7].

$$\overline{f}\ (n_1, n_2)\ {}_{=} \sum_{x_2=0}^{N_2-1} \sum_{x_1=0}^{N_1-1} f(x_1, x_2) e^{-2\pi i \left(\frac{N_1 x_1}{N_1} + \frac{N_2 x_2}{N_2}\right)} \quad \text{.................................} \quad (6)$$

The following function for discrete curvelet transform is not integrated with curvelet coefficients.

$$f(m_1, m_2) = \sum_{j=1}^{j} \sum_{l=0}^{L_j-1} \sum_{k_1=0}^{K_{j,l,1}-1} \sum_{k_2=0}^{K_{j,l,2}-1} c_{jlk} s_{jlk}(x_1, x_2) \quad \text{.......................................} \quad (7)$$

Where k= (k1, k2), s is the curvelet on point j with direction l and spatial shift k.

$$\sum_{jlk} |C_{jlk}|^2 = \sum_{x_1, y_2} + |f(x_1, x_2)|^2 \quad \text{...} \quad (8)$$

In DCT, original signal passes through two filters such as a low-pass filter and a high-pass filter and emerges as two signals. The output of a low pass filter is called as approximation coefficients and the outcome of high-pass filter as detail coefficients. In speech signals, the little frequency elements characterize a signal more than its high frequency elements and thus the low frequency elements h[m] of superior than high frequency signals g[m]. The following high pass and low pass filtering of the signal by the following equations [5].

$$X_{high}[\mathrm{k}] = \sum_m y(m) g(2k - m) \quad \text{..........} \quad (9)$$

$$X_{low}[\mathrm{k}] = \sum_m y(m) h(2k - m) \quad \text{..........} \quad (10)$$

Where X_{high} are the detail coefficients and X_{low} are the approximation coefficients which are the outcomes of the high pass and low pass filters get by sub sampling by factor 2. This filtering event is continued until the necessary level is get according to Mallat algorithm. The DCT decomposition tree and pseudocde is given in figure 3 [5].

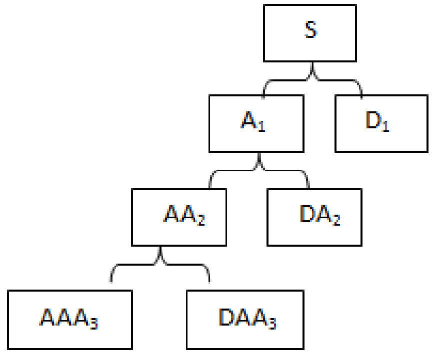

Figure 3. DCT Decomposition Tree

```
function   y = curvelet ( st, R, md )
            if exist( 'md', 'var' )
                      md = 0;
            end
           l = size( st, 1 );
           if md == 0
                   zp = kkt_iso_idwt( st(1,:), R, md );
           else if md == 1
                   zp = kkt_iso_idwt( st{1}, R, md );
          end
          for   ii = 2:l
          if md == 0
               zp (ii,:) = kkt_iso_idwt( rt(ii,:), R, md );
          else if md == 1
               zp (ii,:) = kkt_iso_idwt( st{ii}, R, md );
         end
    end
             xr = rectopolar_2_cart( yp );
             y = i_kkt_2( yr );
     [l1 l2] = size( y );
            y = y(1:l1-1,1:l2-1);
```

The curvelet s is described through its discrete Fourier transform as $\bar{s}jok(m1,m2) = Vj(m1,m2)\, e^{-2\pi i}(k1m1 / Kj0,1 + k2m2 / Kj0,2)$ and $\bar{s}jlk = S^{T}\theta_1 \bar{s}jok$, Here, S_θ is termed as shear matrix, which stands for grid on the curvelet is assessing by an angle Øl. The slopes described by the angles Øl, it is equi-spaced angle.

4. PATTERN CLASSIFICATION AND RECOGNITION

A Hidden Markov Model is a combination of two stochastic processes; the primary is a Markov chain featured by a finite set μt of non observable N states and the transition probabilities, $b_{ij} = Q(s_j^{t+1}|s_i^t)$, $1 \le i, j \le N$, amid them. The second stochastic procedure generate the sequence of T annotations which based on the probability density function of the observation design represented by $aa_j(O_t) = Q(O_t|s_j^t)$. This paper presents, use combination of Gaussian densities. The primary distribution state represented by $\pi_i = Q(s_i^1)$, $1 \le i \le M$. So, HMM λ (B, A, Π) is branded by the transition matrix B = { a_{ij} }, the observation model A = {aj (Ot)} and the preliminary state distribution Π = {πi }.

There exist three basic issues of HMM, for solving these issues using models in real world applications. The primary issue is evaluation problem; it searches to process the probability Q (O/λ), the observation sequence O was produced by the model λ. This probability can be obtained by using forward propagation. Recursively, it estimates the forward variable:

$$\alpha_t(\mathrm{i}) = Q(O_1, O_2, \ldots, O_t, p_t = S_i \mid \lambda) \quad \text{......(11)}$$

$$\alpha_t(\mathrm{i}) = (\textstyle\sum_{i=1}^{N} \alpha_{t-1}(i) b_{ij})\, a_j(O_t) \quad \text{......(12)}$$

Then, Q(O/λ) = $\sum_{i=1}^{N} \alpha_T(\mathrm{i})$ is getting by summation of end point forward variables and the backward propagation can be used to find the solution of this issue. Unlike forward, the backward propagation goes backward. At each instant, it calculates the backward variable:

$$\beta_t(\mathrm{i}) = Q(O_{1+1}, O_{t+1}, \ldots, O_r, q_t = S_i, \lambda) \quad \text{......(13)}$$

$$\beta_t(\mathrm{i}) = (\textstyle\sum_{j=1}^{M} b_{ij}\, \mathrm{aj}\, (O_{t+1})\beta_{t+1} \quad \text{...... (14)}$$

Finally, $Q(O \mid \lambda) = \sum_{i=1}^{N} b_i(\mathrm{i})\beta_t(\mathrm{i})$ is obtained by combining the forward and backward variable. The second problem is named the decoding problem. It searching for predicts state sequence S, it is generated by O. The Viterbi [8] algorithm provides the solution for this issue. It starts from the preliminary instant t = 1, for every moment t, it measures δt(i) for every state i, then it keeps the state which have the maximum δt = maxp1,p2,...,pt−1 Q (p1, p2, . . . pt−1, pt = i, O1O2......Ot-1 | λ = max 1 ≤ i≤ M (δt−1 (i) bij) aj (Ot). When, the algorithm reaches the past instance t = T, it keeps the state with maximum δT. Finally, Viterbi algorithm back-track the sequence of states as the pointer in every instance indicates with t. The current issue is the learning, it seeks to control the model parameters in order to take benefit of Q (O | λ). Baum-Welch [8] method is extensively used for current issue. This algorithm uses for re-estimation of forward and backward variables of model parameters. The principal apparatus of a vast vocabulary working for continuous speech recogniser are described in Figure 4.

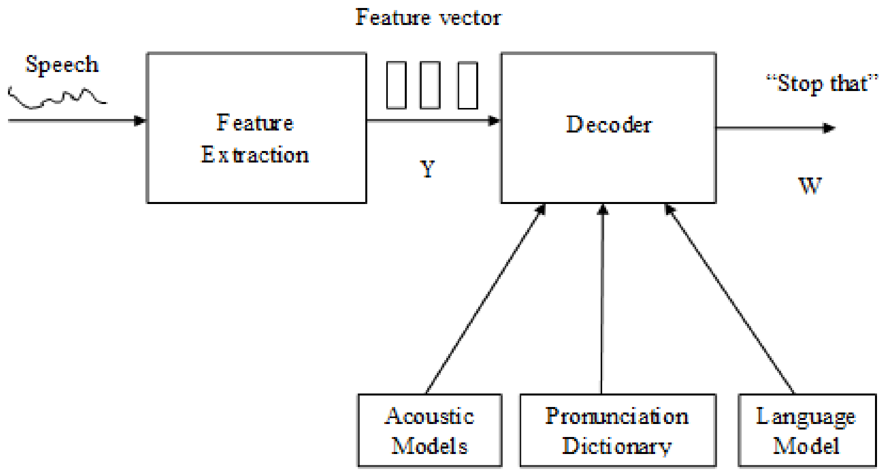

Figure 4. Architecture of HMM-based Recognizer

The given input audio waveform from a microphone is converted into sequence of flat size acoustic vectors Y1:E = y1,...,yE. Then the decoder attempts to discover the sequence of words w1:L = w1,...,wL which is most likely to have generated by Y , i.e. the decoder tries to find the following formula

$$w\hat{} = \arg\max_w \{P(w|Y)\} \quad (15)$$

Since P (w|Y) is not easy to connect model directly, Bayes' Rule is used for transformation (15) into the equivalent problem for finding solution:

$$w\hat{} = \arg\max_w \{p(Y|w)\, P(w)\} \quad (16)$$

The likelihood p(Y |w) is strong-minded by an acoustic model and the prior P(w) is represented by a language model. The basic unit of noise determined by the auditory model is the phone.

The above equations are used for the speech identification process generated by model. Suppose S represented as speech signal in the system, S consists of the path in the syntactic network. The principal stage is to convert S into a sequence of acoustic vectors using the same feature extraction method used for training and we obtain sequence of observations O. The most likely path enhanced the probability of observing O, this model Q (O/ λ). This probability can be discover either by using the forward algorithm or Viterbi algorithm. For any structure employs HMM method three fundamental algorithms which are evaluation, training, and classification algorithms. In classification algorithm, the recognition process is enabled for any unknown utterance by identifying the unidentified observations sequence via choosing the most likely class to have generated the observation sequence. In training algorithm, the model is responsible to store data collected for a specific language. In the evaluation algorithm, the probability of an observation sequence is computed for matching processes. The classification algorithm was employed for a given observations O = O1, O2, O3 … OE. A chosen class was computed using the following equations [10]:

$$\text{Chosen\_Class} = \arg MAX\,[\,P\,(M_{class}|\,O]\quad (17)$$

Therefore, by applying Bayesian rule to find Q (N_{class}| O] the probability was computed as

$$Q(N_{class}|\,O) = \frac{Q(O|N_{class})\,Q\,(N_{class})}{Q\,(O)} \quad (18)$$

In the recognition process, probability of each word was calculated in order to match the occurrence of specific word with another one in the vocabulary table. The elements of the study, these elements sequence are denoted by O_e and elements of the state sequence are denoted by S_e.

Using this structure, we can characterize the system with a set of state transition probabilities a_{ij} = P_r{transaction state i to j} and state dependent observation probabilities $a_j(O_e)$ =$Q_r\{O_e|S_e = j\}$. In the case that the observations come from a discrete and finite set U of size N, we can characterize the conditional probability of observing each element of that set as $b_{jk} = \{O_t = U_k\}$where $U_k = U$. Based on these definitions, an HMM with n states can be characterized by a set of parameters λ = { π, B, A}. Where B_i = [b] is the M x M state transaction matrix, A = [a_{ij}] is the M x N observation matrix, and π = (1,0,…0) is the 1x M initial state vector. In this concern set of parameters represented by λ, these parameters using for forward-backward algorithm as shown in (19), (20) and (21). It can be used to efficiently process the probability of an observation sequence of length E. The forward probabilities and backward probabilities are represented as $\alpha_1(j) = \pi_j a_j(O_e)$and $\beta_t(j) = 1 \forall j$ respectively [10].

$$Q_r\{O|\lambda\} = \sum_{j=1}^{N} \alpha_e(j) \quad \text{.........} (19)$$

$$\alpha_{e+1}(j) = \left[\sum_{j=1}^{M} \alpha_e(i) b_{ij}\right] a_j(O_e) \quad 1 \leq t \leq E - 1 \quad \text{.........} (20)$$

$$\beta_e(i) = \sum_{j=1}^{M} b_{ij}\, a_j(O_e) \beta_{e+1}(j) \quad E - 1 \geq t \geq 1 \quad \text{.........} (21)$$

The Baum-Welch algorithm provide the set of re-estimation formulas shown below (22) and (23), these formulas denotes a novel estimation of that parameter.

$$\bar{b}_{ij} = \frac{\sum_{t=1}^{T-1} \alpha_e(i) b_{ij} a_j(O_e) \beta_{e+1}(j)}{\sum_{e=1}^{T-1} \alpha_e(j) \beta_e(j)} \quad \text{.........} (22)$$

$$\bar{b}_{jk} = \frac{\sum_{t \ni O_e = v_k} \alpha_e(j) \beta_e(j)}{\sum_{e=1}^{T-1} \alpha_t(j) \beta_e(j)} \quad \text{.........} (23)$$

The following algorithm for HMM based IOS used for finding best sequence of hidden states.

```
function   viterbi (observations of  len E, state-graph of len M) returns best-path
        create a path probability matrix viterbi[M+2,E]
        for each state s from 1 to M do      ; initialization step
            Viterbi[s,1] ← a0,s * bs(o1)
            Backpointer[s,1] ← 0
   for each time step t from 2 to E do                    ; recursion step
   for each state s from 1 to M do
        viterbi[s,e] ← max_{s'=1}^{M} viterbi[ s', e-1] * a_{s',s} * bs(oe)
         backpointer[s,t] ←  argmax_{s'=1}^{M}  viterbi[s', e-1] * a_{s',s}
  viterbi[qF,E] ← max_{s=1}^{M} viterbi[s,T] * a_{s,qF}     ;termination step
  backpointer[qF,E] ←  argmax_{s=1}^{M}  viterbi[s, T] * a_{s,qF}
  return the backtrace path by following backpointers
                                           to states back in time from  backpointer[qF,E]
```

Figure 5. HMM algorithm IOS used for finding optimal sequence of hidden states

The function viterbi (observations of len E, state-graph of len M) returns best-path create a path probability matrix viterbi[N+2,T] for each state s from 1 to M do ; initialization step viterbi[s,1]←$a_{0,s} * b_s(o_1)$ backpointer[s,1]←0 for each time step t from 2 to E do ; recursion step for each state s from 1 to M do viterbi[s,e]← M max s_0=1 viterbi[s_0 ,e −1] ∗ $a_{s,0}$,s ∗ $b_s(o_e)$ backpointer[s,e]← Margmax s_0=1 viterbi[s_0 ,t −1] ∗ a_s 0 ,s viterbi[qF ,E]← M max s=1 viterbi [s, E] ∗ a_s,qF ; termination step backpointer[qF ,E]← N argmax s=1 viterbi[s,E] ∗ as,qF ; termination step return the backtrace path by following backpointers to states back in time from backpointer [qF ,E] [9].

5. PERFORMANCE OF THE SYSTEM IN NOISY ENVIRONMENTS

The throughput of speech identification process is often presented in terms of speed and accuracy of the system. Accuracy may be calculated word error rate (WER) and speed is calculated in terms of real time factor [3]. Single Word Error Rate (SWER) and Command Success Rate (CSR) are other measures for accuracy calculation. In this section, we present experiments in order to validate our approach. We use curvelet Coefficients as feature vectors and also a three state HMM, two Gaussian mixtures. The speech recognition schemes are very expensive. Consequently, using the HMM recognizer to a great extent reduce the cost of these systems. In this HMM-based speech recognition system, five audio files such as BJW, JLS, JRM, LPN and RM2 are modelled in HMM. [8] In recognition, additional number of states in HMM, it generate good recognition rate or accuracy. The following tables 1, 2 and 3 present the percentages of recognition rate for speech recognition.

Table 1: Percentage Accuracy for five states of HMM (N=5)

Dataset	**Total test**	**correct rate**	**Error rate**	**Accuracy**
BJW	85	64	21	75.29%
JLS	85	65	20	76.47%
JRM	85	62	23	72.94%
LPN	85	59	26	69.41%
RM2	85	63	22	74.11%

Table 2: Percentage Accuracy for seven states of HMM (N=7)

Dataset	**Total test**	**correct rate**	**Error rate**	**Accuracy**
BJW	85	68	21	80.00%
JLS	85	69	20	81.17%
JRM	85	66	23	77.64%
LPN	85	63	26	74.11%
RM2	85	67	22	78.82%

Table 3: Percentage Accuracy for ten states of HMM (N=10)

Dataset	**Total test**	**correct rate**	**Error rate**	**Accuracy**
BJW	85	72	21	84.70%
JLS	85	70	20	82.35%
JRM	85	69	23	81.17%
LPN	85	66	26	77.64%
RM2	85	71	22	83.52%

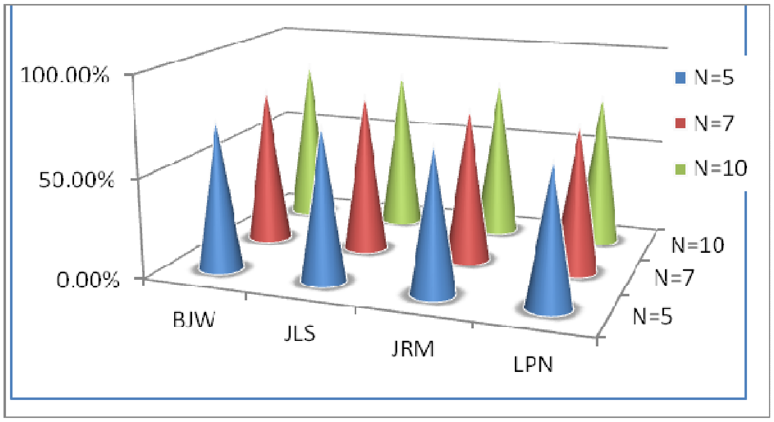

Figure 6. Comparison of Recognition Rates of Speech Recognition System for N=5, 7, 10

The past few decades, extensive research has been carried out on different possible developments of automatic speech recognition (ASR) systems. The most important algorithms in the area of ASR are the curvelet transform with the combination of hidden Markov models. Whatever may be so many other methods also available such as wavelet-based transforms, artificial neural networks with support vector machine classifier, which are becoming more popular, this paper generate a comparative study based on diverse approaches that are proposed for the process of ASR. The following table displays the performance comparison of dissimilar feature classifiers for speech recognition process.

Table 4. Performance comparison of different feature classifier for speech recognition system

Classifier	Precision	Recall	Specificity	F-Measure
Naive Bayesian Approach	94.69%	88.50%	84.20%	0.92%
Support Vector Machine	93.33%	84.80%	84.04%	0.88%
Artificial Neural Network	93.02%	87.00%	86.04%	0.90%
Vector Quantization	92.02%	85.32%	83.01%	0.91%
Gaussian Mixture Model	91.23%	82.32%	84.12%	0.89%
Dynamic Time Wrapping	90.54%	81.54%	85.23%	0.93%
Template Based Approach	92.31%	85.41%	82.12%	0.89%
Pattern Recognition Model	94.52%	80.12%	86.32%	0.94%
Acoustic Model	91.45%	82.12%	84.12%	0.96%
Hidden Markov Method	98.63%	90.12%	83.14%	0.95%

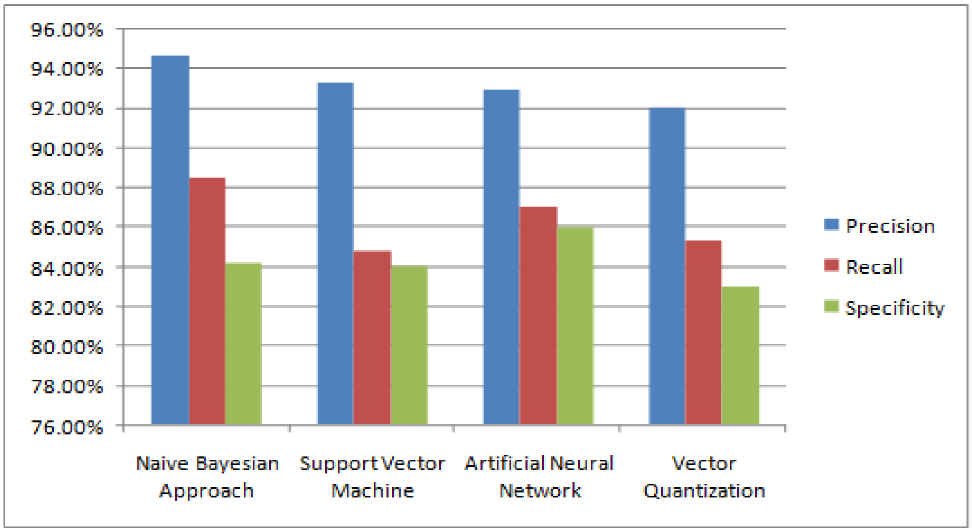

Figure 7. Performance comparison of dissimilar feature classifier for speech recognition process

The table given below shows the comparison of various classifiers

Table 5. Comparison of various classifiers

S. No	Classifier	Recognition Accuracy (%)
1	Naive Bayesian Approach	84.13
2	Support Vector Machine	88.25
3	Artificial Neural Network	90.23
4	Vector Quantization	90.14
5	Gaussian Mixture Model	90.32
6	Dynamic Time Wrapping	91.45
7	Template Based Approach	91.13
8	Pattern Recognition Model	93.25
9	Acoustic Model	92.35
10	Hidden Markov Method	97.72

6. CONCLUSION

Now a day's biggest challenge is speech recognition on a daily basis, in which deformation and noise in the environment are present and hold back recognition process. A speech recognition system contains four stages: Analysis, Feature Extraction, and Modelling and Matching techniques. In the field of speech recognition, lot of research done but still speech recognition systems are not a hundred percent accurate, all with its own benefits and drawbacks. Whatever it is accuracy for speech recognition still concentration for disparity of context, speaker's variability, and environment conditions. In this paper, we developed curvelet based Feature Extraction method (CFE) for speech detection in noisy background and the input speech signal is decomposed into dissimilar frequency channels using characteristics of curvelet transform. Feature extraction is a vital stage in speech processing using curvelet transforms, the computational complexity and the feature vector size are successfully reduced. They have better accuracy, varying window size because of which they are suitable for non –stationary signals. Thus curvelet transform is an elegant tool for the analysis of non-stationary signals like speech. Hence various feature extraction methods can be used for dissimilar kinds of applications. Though there are various common applications between these feature extraction techniques, among all these techniques curvelet transform gives better system accuracy. Discrete hidden markov model can be used for classification for better word recognition as they consider time distribution of speech signals. The different approaches developing for speech recognition system are compared with their merits and demerits. The HMM classification method achieved the utmost accuracy in terms of identification rate for conversational with 80.1%, scientific phrases with 86%, and control with 63.8 % recognition rates. The objective of this paper is to epitomize the CFE method and classification stages in speech recognition system. The statistical results shows that signal recognition accuracy will be increased by using discrete Curvelet transforms over conventional methods.

REFERENCES

[1] A. Madan and Divya Gupta," Speech Feature Extraction and Classification: A Comparative Review" International Journal of Computer Applications (0975 – 8887) Volume 90 – No 9, March 2014.

[2] Nidhi Desai, Prof.Kinnal Dhameliya and Prof.Vijayendra Desai," Feature Extraction and Classification Techniques for Speech Recognition: A Review", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 3, Issue 12, December 2013.

[3] Nidhi Srivastava and Dr.Harsh Dev"Speech Recognition using MFCC and Neural Networks", International Journal of Modern Engineering Research (IJMER), March 2007.

[4] Bishnu Prasad Das and Ranjan Parekh, "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, May-June 2012.

[5] Sonia Sunny, David Peter S, K Poulose Jacob, "Design of a Novel Hybrid Algorithm for Improved Speech Recognition with Support vector Machines Classifier", International Journal of Emerging Technology and Advanced Engineering, vol.3, pp.249-254, June 2013.

[6] M. C. Shrotriya, Omar Farooq, and Z. A. Abbasi, " Hybrid Wavelet based Lpc Features for Hindi Speech Recognition ", International Journal of Information and Communication Technology, March 2008.

[7] subhani shaik and shaik subhani,"Detection and classification of power quality disturbances using curvelet transform and support vector machine", International Conference on Information Communication and Embedded Systems (ICICES 2016). Pages: 1 - 8, DOI: 10.1109/ICICES.2016.7518870.

[8] Su Myat Mon, Hla Myo Tun," Speech-To-Text Conversion (STT) System Using Hidden Markov Model (hmm)", international journal of scientific & technology research volume 4, issue 06, june 2015 issn 2277-8616.

[9] Speech and Language Processing. Daniel Jurafsky & James H. Martin, Draft of August 7

AUTHORS

Mr. N srinivasaro received his B.Tech degree from JNTU Hyderabad, received his M.Tech degree from Acharya Nagarjuna University, Guntur, A.P. His area of interests are Image processing, Computer Networks and Data Mining.

Ms. Ch. Anuradha received her B.Tech degree from Acharya Nagarjuna University, Guntur and received her M.Tech degree from JNTUK, Kakinada. She has 4 years of Industry experience and 5 years of techning experinece. She published 8 papers in National and Inter national journals. Her area of Interests are Image processing, Data Mining, Network Security.

Dr. SVN Naga Srinivasu received his Phd from Acharya Nagarjuna University, Guntur, Andhara Pradesh. Present he is working as a research supervisor at Nagarjuna University, Guntur. He published more number of papers in National and Inter National Journals. His area of Interests are Image Processing, Computer Networks, Software Engineering and Data Mining.