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

Nidamanuru Srinivasa Rao¹, Chinta Anuradha², Dr. SV Naga Sreenivasu³

¹Scholar, Department of CSE, Acharya Nagarjuna University, Guntur, AP
²Asst.Prof, Department of CSE, VR Siddhartha Engineering College, Vijayawada, AP
³Professor, 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 phase 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](image)

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

\[
x(e) = u(e) + y(e) \text{ .................................................. (1)}
\]
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 $O = o_1o_2 \cdots$ be a sequence of parameters vectors, where $o_t$ is observed at time $t \in \{1 \ldots \tau\}$. Given a dictionary $D$ with words $w_i \in D$, the recognition problem is

$$W^* = \arg \max_W P(W|O)$$

Where $W^*$ is the recognised word, $P$ is the probability measure and $W = w_1 \ldots w_k$ a word sequence and transform (1) in a suitable calculable form as:

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

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 $P(O|w_i)$. On Simplification (1) becomes

$$W^* = \arg \max_W P(W|O)P(W)^{s|W|^p}$$

In the log domain, the total likelihood is calculated as

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

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 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 recognizes 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 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. Among 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 \( f(x_1, x_2) \), be function \( y_1 = 0, 1, \ldots, N1 - 1, x_2 = 0, 1, \ldots, N2 - 1 \), based on this function, we define the discrete Fourier transform [7].

\[
\overline{f}(n_1, n_2) = \sum_{x_1=0}^{N1-1} \sum_{x_2=0}^{N2-1} f(x_1, x_2) e^{-2\pi i \left( \frac{n_1 x_1}{N1} + \frac{n_2 x_2}{N2} \right)} \tag{6}
\]

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

\[
f(m_1, m_2) = \sum_{j=1}^{J} \sum_{l=0}^{L1-1} \sum_{k=0}^{K1-1} \sum_{s=0}^{S1-1} c_{jls} \delta(x_1, x_2) \tag{7}
\]

Where \( k = (k_1, k_2) \), \( s \) is the curvelet on point \( j \) with direction \( l \) and spatial shift \( k \).

\[
\sum_{jls} |c_{jls}|^2 = \sum_{x_1, x_2} |f(x_1, x_2)|^2 \tag{8}
\]

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].

\[
X_{\text{high}}[k] = \sum_{m} y(m) g(2k - m) \tag{9}
\]

\[
X_{\text{low}}[k] = \sum_{m} y(m) h(2k - m) \tag{10}
\]

Where \( X_{\text{high}} \) are the detail coefficients and \( X_{\text{low}} \) 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 pseudocode is given in figure 3 [5].
function \( y = \text{curvelet}( \text{st}, R, \text{md} ) \)

\[
\begin{align*}
\text{if exist('md', 'var')} & \\
\quad \text{md} &= 0; \\
\text{end} \\
\text{l} &= \text{size(st, 1)}; \\
\text{if md == 0} & \\
\quad \text{zp} &= \text{kkt_iso_idwt}( \text{st(1,:)}, R, \text{md} ); \\
\text{else if md == 1} & \\
\quad \text{zp} &= \text{kkt_iso_idwt}( \text{st{1}}, R, \text{md} ); \\
\text{end} \\
\text{for ii = 2:l} & \\
\text{if md == 0} & \\
\quad \text{zp (ii,:) = kkt_iso_idwt( rt(ii,:), R, md );} \\
\text{else if md == 1} & \\
\quad \text{zp (ii,:) = kkt_iso_idwt( st{ii}, R, md );} \\
\text{end} \\
\text{end} \\
\text{xr = rectopolar_2_cart( yp );} \\
\text{y = i_kkt_2( yr );} \\
[l_1, l_2] &= \text{size}(y); \\
\text{y} &= y(1:l_1,1:l_2); \\
\end{align*}
\]

The curvelet \( s \) is described through its discrete Fourier transforms as \( \tilde{s}(\mathbf{m}_1, \mathbf{m}_2) = V j(\mathbf{m}_1, \mathbf{m}_2) e^{-2\pi i (k_1 m_1 / K_1 + k_2 m_2 / K_2)} \) and \( \hat{s}(\mathbf{k}) = S^T \hat{\theta}_{\mathbf{k}} \hat{s}_{\mathbf{k}} \). Here, \( S_0 \) is termed as shear matrix, which stands for grid on the curvelet is assessing by an angle \( \Theta_0 \). The slopes described by the angles \( \Theta_0 \), 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 of non-observable $S$ states and the transition probabilities, $b_{ij} = Q(s_{i+1}|s_i), 1 \leq i, j \leq N$, 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 $a_{i}(O_t) = Q(O_t|s_i^t)$. This paper presents, use combination of Gaussian densities. The primary distribution state represented by $\pi_i = Q(s_i^1), 1 \leq i \leq M$. So, HMM $\lambda(B, A, \Pi)$ is branded by the transition matrix $B = \{a_{ij}\}$, the observation model $A = \{a_{i}(O_t)\}$ and the preliminary state distribution $\Pi = \{\pi_i\}$.

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 $Q(O/\lambda)$, the observation sequence $O$ was produced by the model $\lambda$. This probability can be obtained by using forward propagation. Recursively, it estimates the forward variable:

$$\alpha_t(i) = Q(O_1, O_2, ..., O_t, p_t = S_t | \lambda)$$

$$\alpha_t(i) = (\sum_{i=1}^{N} \alpha_{t-1}(i)b_{ij}) a_{i}(O_t)$$

Then, $Q(O/\lambda) = \sum_{i=1}^{N} \alpha_T(i)$ 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:

$$\beta_t(i) = Q(O_{t+1}, O_{t+2}, ..., O_T, q_t = S_t, \lambda)$$

$$\beta_t(i) = (\sum_{i=1}^{N} b_{ij} a_{j}(O_{t+1})\beta_{t+1})$$

Finally, $Q(O/\lambda) = \sum_{i=1}^{N} b_{1i} \beta_1(i)$ 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 $\delta_t(i)$ for every state $i$, then it keeps the state which have the maximum $\delta_t = \max_{1 \leq i \leq M} Q(p_1, p_2, ... p_{t-1}, pt = i, O_1O_2...O_{t-1} | \lambda = \max 1 \leq i \leq M (\delta_{t-1}(i) bij) a_{ij}(O_t)$. When, the algorithm reaches the past instance $t = T$, it keeps the state with maximum $\delta_T$. 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 $Q(O/\lambda)$. 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.
The given input audio waveform from a microphone is converted into a sequence of flat-size acoustic vectors $Y_{1:E} = y_1, ..., y_E$. Then the decoder attempts to discover the sequence of words $w_{1:L} = w_1, ..., w_L$ which is most likely to have generated by $Y$, i.e., the decoder tries to find the following formula

$$ w^* = \arg \max_w \{ P(w|Y) \} \quad \ldots \quad (15) $$

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:

$$ w^* = \arg \max_w \{ p(Y|w) P(w) \} \quad \ldots \quad (16) $$

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 $Q(O/\lambda)$. 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 $O = O_1, O_2, O_3 ... O_E$. A chosen class was computed using the following equations [10]:

$$ \text{Chosen} \_\text{Class} = \arg \max_{M_{\text{class}}} \{ P(M_{\text{class}}|O) \} \quad \ldots \quad (17) $$

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

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

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 $S_e$. 

![Figure 4. Architecture of HMM-based Recognizer](image-url)
Using this structure, we can characterize the system with a set of state transition probabilities \( a_{ij} = P\{\text{transaction state } i \text{ to } j\} \) and state dependent observation probabilities \( a_j(0_e) = Q\{0_e | S_e = j\} \). 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 \( b_{ik} = P\{0_e = u_k | S_e = j\} \). Based on these definitions, an HMM with \( n \) states can be characterized by a set of parameters \( \lambda = \{ \pi, B, A \} \). Where \( B = [b] \) is the \( M \times M \) state transaction matrix, \( A = [a_{ij}] \) is the \( M \times N \) observation matrix, and \( \pi = (1,0,\ldots,0) \) is the \( 1 \times M \) initial state vector. In this concern set of parameters represented by \( \lambda \), 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 \( \alpha_q(j) = \pi_i a_i(0_e) \) and \( \beta_q(j) = 1^{\forall j} \) respectively [10].

\[ \alpha_{t+1}(j) = \left[ \sum_{j=1}^{M} a_j(i) b_j(0_e) \right] \alpha_t(0_e) \quad 1 \leq t \leq E - 1 \]  

\[ \beta_t(j) = \sum_{j=1}^{M} b_{ij} a_j(0_e) \beta_{t+1}(j) \quad E - 1 \geq t \geq 1 \]

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.

\[ \bar{B}_{ik} = \frac{\sum_{t=1}^{E} a_{t+1}(i) b_{ij} a_j(0_e) \beta_{t+1}(j)}{\sum_{t=1}^{E} a_{t+1}(i) \beta_{t+1}(j)} \]  

\[ \bar{B}_{ik} = \frac{\sum_{t=1}^{E} a_{t+1}(i) b_{ij} a_j(0_e) \beta_{t+1}(j)}{\sum_{t=1}^{E} a_{t+1}(i) \beta_{t+1}(j)} \]

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

```plaintext
function viterbi(observations of len E, state-graph of len M) returns best-path
create a path probability matrix viterbi[N+2,E]
for each state s from 1 to M do  ; initialization step
    viterbi[s,1] ← a_0,s * b_s(0_e)
    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=1}^M viterbi[s',e-1] * a_{s',s} * b_s(0_e)
        backpointer[s,e] ← argmax_{s'-1=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 backrace 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.
E] ∗ aₙ,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)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Total test</th>
<th>Correct rate</th>
<th>Error rate</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BJW</td>
<td>85</td>
<td>64</td>
<td>21</td>
<td>75.29%</td>
</tr>
<tr>
<td>JLS</td>
<td>85</td>
<td>65</td>
<td>20</td>
<td>76.47%</td>
</tr>
<tr>
<td>JRM</td>
<td>85</td>
<td>62</td>
<td>23</td>
<td>72.94%</td>
</tr>
<tr>
<td>LPN</td>
<td>85</td>
<td>59</td>
<td>26</td>
<td>69.41%</td>
</tr>
<tr>
<td>RM2</td>
<td>85</td>
<td>63</td>
<td>22</td>
<td>74.11%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Total test</th>
<th>Correct rate</th>
<th>Error rate</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BJW</td>
<td>85</td>
<td>68</td>
<td>21</td>
<td>80.00%</td>
</tr>
<tr>
<td>JLS</td>
<td>85</td>
<td>69</td>
<td>20</td>
<td>81.17%</td>
</tr>
<tr>
<td>JRM</td>
<td>85</td>
<td>66</td>
<td>23</td>
<td>77.64%</td>
</tr>
<tr>
<td>LPN</td>
<td>85</td>
<td>63</td>
<td>26</td>
<td>74.11%</td>
</tr>
<tr>
<td>RM2</td>
<td>85</td>
<td>67</td>
<td>22</td>
<td>78.82%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Total test</th>
<th>Correct rate</th>
<th>Error rate</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BJW</td>
<td>85</td>
<td>72</td>
<td>21</td>
<td>84.70%</td>
</tr>
<tr>
<td>JLS</td>
<td>85</td>
<td>70</td>
<td>20</td>
<td>82.35%</td>
</tr>
<tr>
<td>JRM</td>
<td>85</td>
<td>69</td>
<td>23</td>
<td>81.17%</td>
</tr>
<tr>
<td>LPN</td>
<td>85</td>
<td>66</td>
<td>26</td>
<td>77.64%</td>
</tr>
<tr>
<td>RM2</td>
<td>85</td>
<td>71</td>
<td>22</td>
<td>83.52%</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Precision</th>
<th>Recall</th>
<th>Specificity</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayesian Approach</td>
<td>94.69%</td>
<td>88.50%</td>
<td>84.20%</td>
<td>0.92%</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>93.33%</td>
<td>84.80%</td>
<td>84.04%</td>
<td>0.88%</td>
</tr>
<tr>
<td>Artificial Neural Network</td>
<td>93.02%</td>
<td>87.00%</td>
<td>86.04%</td>
<td>0.90%</td>
</tr>
<tr>
<td>Vector Quantization</td>
<td>92.02%</td>
<td>85.32%</td>
<td>83.01%</td>
<td>0.91%</td>
</tr>
<tr>
<td>Gaussian Mixture Model</td>
<td>91.23%</td>
<td>82.32%</td>
<td>84.12%</td>
<td>0.89%</td>
</tr>
<tr>
<td>Dynamic Time Wrapping</td>
<td>90.54%</td>
<td>81.54%</td>
<td>85.23%</td>
<td>0.93%</td>
</tr>
<tr>
<td>Template Based Approach</td>
<td>92.31%</td>
<td>85.41%</td>
<td>82.12%</td>
<td>0.89%</td>
</tr>
<tr>
<td>Pattern Recognition Model</td>
<td>94.52%</td>
<td>80.12%</td>
<td>86.32%</td>
<td>0.94%</td>
</tr>
<tr>
<td>Acoustic Model</td>
<td>91.45%</td>
<td>82.12%</td>
<td>84.12%</td>
<td>0.96%</td>
</tr>
<tr>
<td>Hidden Markov Method</td>
<td>98.63%</td>
<td>90.12%</td>
<td>83.14%</td>
<td>0.95%</td>
</tr>
</tbody>
</table>
The table given below shows the comparison of various classifiers

Table 5. Comparison of various classifiers

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

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


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 teaching experience. She published 8 papers in National and International journals. Her area of interests are Image Processing, Data Mining, Network Security.

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