Indian Journal of Science and Technology, *Vol 9(47), DOI: 10.17485/ijst/2016/v9i47/101744, December 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

A HMM Integrated SVM Model for Hindi Speech Recognition

Bhavneesh Sohal and Sandeep Kaur

Computer Science and Engineering, Lovely Professional University, Phagwara, India;
Sohal.16042@lpu.co.in, Sandeep.16827@lpu.co.in

Abstract

Speech is one of the most useful and effective communication medium used as the biometric feature of human being. Speech Recognition is considered as important part of various applications for biometric recognition and translation. Speech can be acquired inexpensively using some mic or phone device. But because of different noise factors that can be included during acquisition increases the complexity in recognition process. Because of this, there is the requirement of more effective and reliable approach for speech recognition. In this present work, a statistical analysis based predictive model is defined to improve the speech recognition. The presented work is defined in the form of a layered model. In first layer of this model, the speech signal improvement is done. To achieve this, the Discrete Wavelet Transform (DWT) and spectral subtraction based approach is defined for noise reduction. This layer improves the speech features. In second layer, the Hidden Markov Model (HMM) improved statistical model is defined to generate the speech features. This layer uses the segmented mean value, standard deviation, variation analysis as the statistical features. In final stage, Distance based mapping is applied on all these features collectively to perform the recognition. The work is implemented for Hindi word recognition. The work is implemented in Matlab environment.

Keywords: Discrete Wavelet Transform (DWT), Hidden Markov Model (HMM), Speech Recognition, Spectral Subtraction, Support Vector Machine (SVM)

1. Introduction

Speech is the form of sound used to provide the verbal communication between human beings. Speech Recognition is a trending research topic because it is one of the most important factors that influence speech recognition performance. The speech recognition is having the significance in different application areas. These applications include the authentication systems, sound reformation, converters etc. Speech is considered as the most critical biometric feature which is processed in complex way. The processing on speech itself is associated with number of problems. These problems are defined under the interaction specification, recognition system and verification system. The conversion of speech to other forms is also done to achieve the effective communication. This kind of communication mechanism also provides the robustness in different relative aspects. These aspects also provide the interaction in different communication forms so that the adaptive speech recognition and processing can be achieved.

Speech signal when captured in real time using some capturing device suffers from number of associated problems or impurities. To achieve the accurate speech recognition and effective speech processing, it is required to identify these problems accurately as well as provide the relative solution so that the accurate speech processing outcome is achieved from the system. The background noise is another vector that increases the signal noise. The background noise can be of fan, traffic or the speech of some other person. The device adaptive impurities also disrupt the speech signal. The variable adaptive communication is here performed so that the recognition process can be optimized.

The robustness of the speech recognition system is required to resolve under the defined issues. The speaker

*Author for correspondence

recognition system under speech variation is required to process. The main cause of speech variation analysis is required to classify the signal under speaker and the recording environment. It is not easy to generate such an environment that will not include any kind of noise in the speech signal. The speech noise can be additive or the multiplicative. The additive noise is included over the speech in fixed frequency throughput the signal. Such noise is easy to observe and discard but the multiplicative noise becomes the part of speech signal and internally disrupts the signal. It is not easy to provide the solution for such speech signals.

In this work, our main objective is to define a wavelet and spectral subtraction approach for speech signal improvement and extracting multiple static features based on HMM approach. Further, system is developed for our national language Hindi as there has been lot of research in area of speech recognition for foreign languages but very few work has been reported in Hindi language. Speech features are extracted to recognize Hindi speech.

2. Related Work

This section discusses the work defined by the earlier researchers.

A.M. Peinado[1] has provided a work on feature extraction using vector quantization approach and the speech signal recognition is done under predictive method. In this work, Hidden Markov Model is defined for speech signal recognition. Author defined the multi model based semi continuous model for speech signal integration to do signal quantization. Author reduced the integrated computation so that more effective and accurate recognition is done. Author applied the SCMVQ modelling method to derive accurate recognition.

Tatsuhiko Kinjo[2] has defined a HMM based approach to improve the speech signal and to improve the recognition model. Author defined the statistical measure so that the speech transformation could be performed. The speech features are extracted to that the accurate signal form is obtained. The speech spectrum analysis and the signal derivation is done to divide the signal under frequency spectrum. Once the signal features are extracted, the predictive model is applied to perform the recognition. The distance analysis is here defined to perform the speech recognition. The result analysis defined by the author shows that the model has provided significant results.

Cong-Thanh Do[3] has provided a work on speech signal processing under spectral method so that the signal error will be reduced. Author used the MFCC and HMM based speech signal processing model so that the signal improvement will be done and the effective signal processing will be obtained. Author provided the work under HMM based model and used it as the predictive model so that the recognition will be improved.

Panikos Heracleous[4] has provided a work on HMM based speech signal processing so that the recognition of the speech signal will be improved. Author defined the Non-Audible Murmur (NAM) speech signal processing model under distance adaptive method so that the predictive signal results will be obtained. Author defined the spectral space based model to improve the signal results and to achieve the significant recognition rate.

A. Revathi[5] has provided a work on speaker dependent model for speech signal recognition. Author provided the digit recognition model using HMM integrated vector quantization model. System worked on isolated digits so that the continuous speech processing would be done. System provided the speech feature generation under isolated digit processing and continuous speech generation. Author improved the recognition rate under vector quantization method.

Jinyu Li[6] has provided a work on wide band method under speech signal processing and bandwidth processing under training method. System used the speech recognition model under signal level derivation and feature extraction. Author defined the flexibility under signal processing and log filtration so that the accurate speech recognition would be done. Author reduced the problem associated with missing feature problem and process on narrow band of speech signal so that the accurate signal processing will be done.

Prof. Ashok Shigli[7] has provided a work on spectral signal processing model under speech processing using HMM based approach. System was developed using MFCC method and provided the spectral signal processing under signal modelling method. Author provided the speech signal modelling under spectral derivation so that the speech signal processing will be done and the effective signal processing and the signal derivation will be obtained from the work.

Mohit Dua and R.K. Aggarwal[8] implemented a Punjabi speech recognition system using Hidden markov model ToolKit (HTK). System used statistical approach called HMM. System was developed using MFCC method.

Preet Saini[9] has provided a work on HMM based speech recognizers so that overall accuracy of system will be improved. System recognized the isolated words using acoustic word model.

Sunija A.P[10] implemented a speech recognition system that recognizes Malayalam language dialects. Only two features were extracted i.e. energy and pitch. System is trained using input feature vector data using information related to known patterns and then they are tested using the test data set.

Rajisha T.M[11] provided a work on MFCC, STE, and pitch as feature extraction techniques. Pattern classification was performed using two classifiers, namely Artificial Neural Network (ANN) and Support Vector Machine (SVM).

Devin Hoesen[12] presented a work on Indonesian speech recognizer. The recognizer is based on the Gaussian Mixture and Hidden Markov Models. System is trained on dictated and spontaneous speech. He proved that Maximum A Posteriori (MAP) and Maximum Mutual Information (MMI) adaption approaches help to increase word accuracy rates as compared to un-adapted approaches.

Our paper aims to design and implement a Hindi speech recognition system using HMM integrated SVM model.

3. Methodology and Implementation

3.1 Problem Definition

The biometric authentication is one the major requirement to accept the validity of a user. There are number of biometric features used to identify a person. One of such biometric feature is speech. Speech is not only used for authentication but also having the significance in many other application such as language translation, text form conversion etc. To use it in different application, the first requirement is to recognize or classify the speech contents. Classification of speech based on contents or topic is always a challenge. The challenges include the various noise vectors including the instrumentation noise, background noise etc. Other challenge is the various in the person pitch when same sentence is spoken by same person. In this paper, the improved classification method is proposed based on HMM improved SVM approach. The work is here presented in three main layers. In first layer, the speech improvement is performed to adjust the signal frequency and to remove the signal noise. This signal enhancement is here performed using wavelet and spectral subtraction method. In second layer, the signal features are extracted using statistical measures and HMM approach. In this phase, multiple statistical features are taken based on fix segment division. In third layer, the SVM is applied to improve the recognition process. The work is implemented in Matlab environment. The work is experimented on Hindi voice signal dataset.

3.2 Objectives

The objectives associated with presented work are given here under

- The main objective of the work is to define a HMM improved SVM approach for Hindi speech recognition.
- The objective of the work is to define a wavelet and spectral subtraction approach for speech signal improvement.
- The objective of the work is to extract the multiple statistical features based on HMM approach.
- The objective of work is to perform the speech signal classification using SVM approach.
- The objective of the work is to improve the recognition rate.

3.3 Significance of Work

The significance of work is defined here under:

- The work is here defined as robustness method that will work effectively against noisy signal.
- The statistical method will improve the reliability of recognition process.

3.4 Research Methodology

In this present work, a layered model is defined to improve the Hindi word recognition against the noisy speech signal. Figure 1 show the layered model defined in this work.

Present work is defined as the three stage model. In first stage of this model, the DWT and spectral subtraction approach is applied to perform the signal filtration. The DWT is used to extract the signal coefficient and map the frequency of signal. Spectral subtraction is the deductive approach used to remove the signal noise. In second stage, the signal feature extraction is performed using HMM based approach. The signal feature taken

here applied based on the signal segment division and the features are extracted on each segment. These features include mean value, standard deviation, peak value, MSE etc. All these feature segments are defined in next sub section. Once the feature set is obtained, the classification of signal is performed using SVM approach. To perform classification, signal set is divided in two subsets called training set and testing set. The process model of this work is shown in Figure 2.

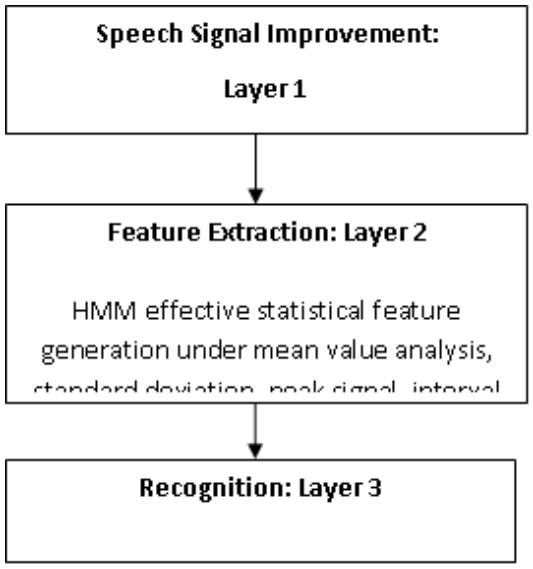

Figure 1. Proposed model

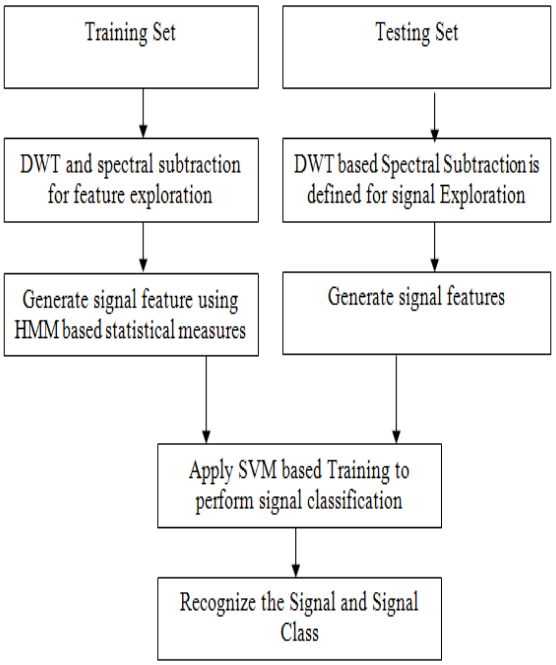

Figure 2. Process model

3.5 Spectral Subtraction Method

The presented work is defined to improve the speech signal using spectral subtraction method. The work is defined to improve the speech signal using high level classification model. Here, the spectral subtraction is applied to improve the signal under noise vector. This method is effective to remove the static noise over the speech signal. The work is defined to analyze the speech signal under magnitude level and perform the adjustment by changing the phase. The adjustment of signal is done using discrete Fourier transformation. The magnitude is done to adjust the noise so that the signal noise suppression is obtained.

The speech signal is represented as signal (n). Here, signal is represented as the speech signal vector and n is identified as the length of the speech signal. The noise is included in the speech so that the window based method is applied over it to remove the noise.

3.6 DWT

The DWT is the decomposition approach used to extract the effective signal features and to maintain the signal effectiveness. In this work, DWT is applied on filtered speech signal to explore the signal features. Here, two-level DWT is applied using sym6 function. The function divides the signal in High and low frequency bands. The information preserving signal exploration is performed using sym6. The basic model of DWT process is shown in Figure 3.

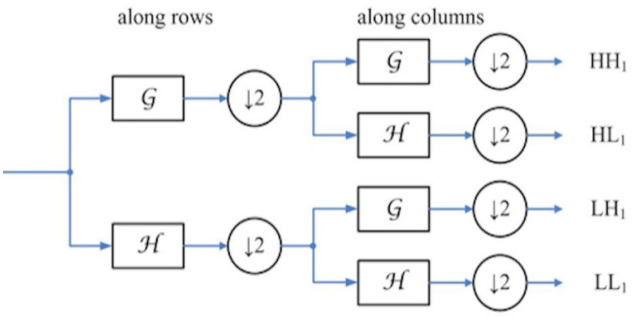

Figure 3. DWT model

Figure 3 shows the DWT model of feature level decomposition and feature extraction. Each time the DWT function is applied, the decomposition coefficients are obtained in terms of high and low frequency bands. Each band can be processed in next level DWT in same way. The process continues till the specific number of DWT levels. The Figure shows the possible generation of DWT features bands after three level decomposition.

3.7 Signal Features

Once the signal features are explored, the next work is defined to extract the signal features. In this work, different types of features are collected to generate the feature set. Based on this feature set, the actual recognition and classification process is performed. The feature extraction is defined under specification of fix segments. The features collected to perform the effective signal analysis are given here under:

a) Segmented-Mean

In this feature, the filtered signal is divided in smaller sub segments with the specification of fix interval or the range. For each range, the frequency level mean is obtained. This mean value is considered as the segmented Mean feature. The feature extraction process is shown in equation (1) and (2).

Segment = GetSegment(Signal, pos, size) (1)

Segmented_Mean=Sum(Freq(Segment))/Length(Segment) (2)

b) Segmented-STD

In this feature, the filtered signal is divided in smaller sub segments with the specification of fix interval or the range. For each range, the frequency level standard deviation is obtained. This standard deviation value is considered as the segmented STD feature. The feature extraction process is shown in equation (3) and (4).

Segment = GetSegment(Signal, pos, size) (3)

Segmented_Mean = Std(Segment) (4)

c) Segmented-Peak

In this feature, the filtered signal is divided in smaller sub segments with the specification of fix interval or the range. For each range, the peak value is obtained. This peak value is considered as the segmented Peak feature. The feature extraction process is shown in equation (5) and (6).

Segment = GetSegment(Signal, pos, size) (5)

Segmented_Peak = Max(Segment) (6)

d) Segmented-Var

The variance is able to identify the changeability in the signal segment. The signal is divided in smaller segment. For this segmented signal, the variance value over the signal is obtained. This segmented value is considered as the effective signal feature. Matlab provide the variance function to extract this feature as shown in equation (7) and (8).

Segment = GetSegment(Signal, pos, size) (7)

Segmented_Var = Var(Segment) (8)

e) Segmented-MSE

MSE is described as the mean square error. It is defined as the quality measure for a signal. Lower the MSE value, more effective the signal is considered. In this work, MSE is considered as the feature vector. The signal is divided in sub segments of fix size and for each segment MSE is computed. The MSE extraction process from the signal segment is shown in equation (9) and (10).

Segment = GetSegment(Signal, pos, size) (9)

Segmented_MSE = Mse(Segment) (10)

f) Segmented-MAE

MAE is described as the mean absolute error. It is defined as the quality measure for a signal. Lower the MAE value, more effective the signal is considered. In this work, MAE is considered as the feature vector. The signal is divided in sub segments of fix size and for each segment MAE is computed. The MAE extraction process from the signal segment is shown in equation (11) and (12).

Segment = GetSegment(Signal, pos, size) (11)

Segmented_MAE = Mae(Segment) (12)

3.8 SVM

In this work, SVM classifier is used to identify the Hindi speech classes. SVM is Support Vector Machine that works with high dimensional data. It is used as a classifier for the identification of classes present in the speech and gives high quality results in the process of classification. It works under the approach of kernel based algorithm. In this method, feature space is computed by identifying the data dependency. SVM is classified along with the kernel function. It also provides the fixed dimensional vector representation. This method helps in identifying the data criticality by working on dissimilar decision functions and data values. Under environmental specifications, data usage based analysis can also be obtained. The simplest form of SVM that can be applied on balanced dataset is known as linear classifier.

4. Experimental Results

4.1 Tool: Matlab

Matlab is defined as a language as well as a tool that provides the rich set of tools to apply different algorithmic

works on different data types. This tool is having the integration with different languages including C, C++ etc. It provides the image processing under matrix formation approach. It provides the command based interface as well as information processing interface to work with image features in effective way. In this work, Matlab is used as the integrated tool.

4.2 Results

4.2.1 Example 1: Normal Signal

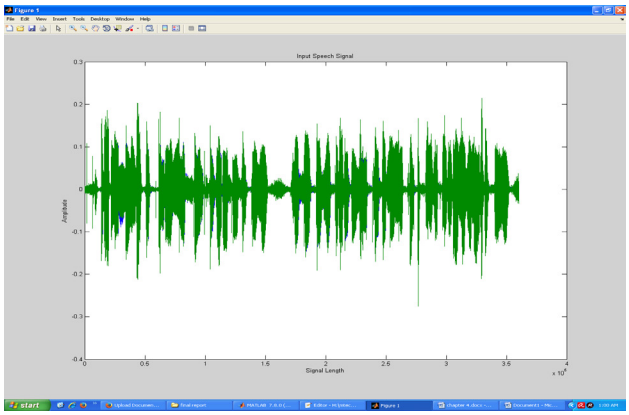

Figure 4. Raw speech signal

Here Figure 4 is showing the raw form of Speech signal. Here x axis represents the signal length and y axis represents the respective amplitude value.

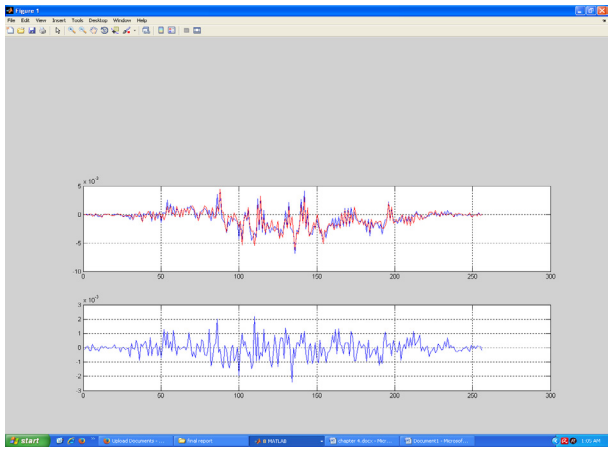

Figure 5. Result of filtration stage

Here Figure 5 is showing the results of filtration stage. The high pass filter is here applied to remove the signal noise. This stage basically removes the baseline drift noise from the signal. As the noise is removed from the signal, some delay is included in the signal itself.

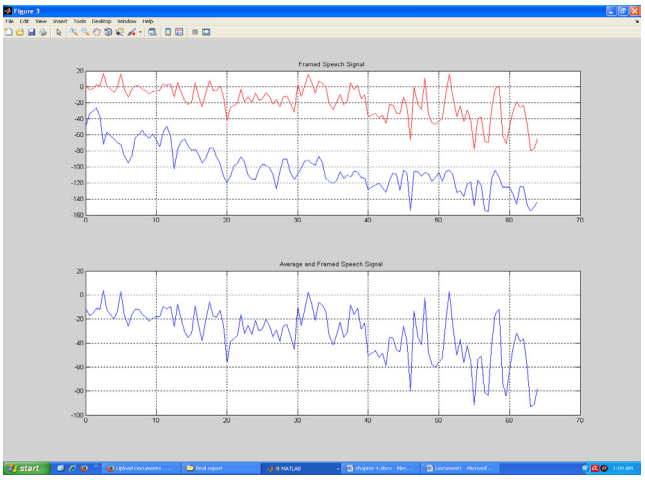

Figure 6. Frame adaptive signal analysis

Here Figure 6 is showing the mapping frame adaptive signal analysis phase so that the noise over the speech signal is removed. The first sub plot is defining the frame adaptation over the signal and the second is showing the mean signal value obtained after applying the frame adaptation.

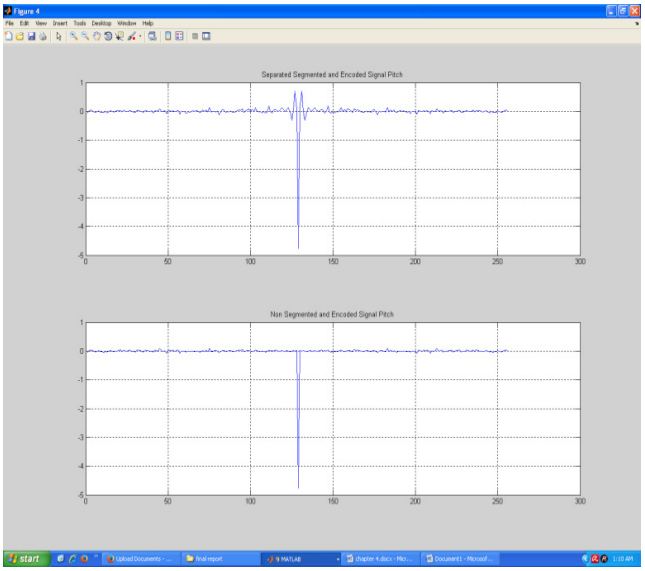

Figure 7. Separated segmented speech signal

Here Figure 7 is showing the separated segmented speech signal obtained using the segmented pitch analysis. The maximum pitch value is obtained to represent the effective signal value analysis.

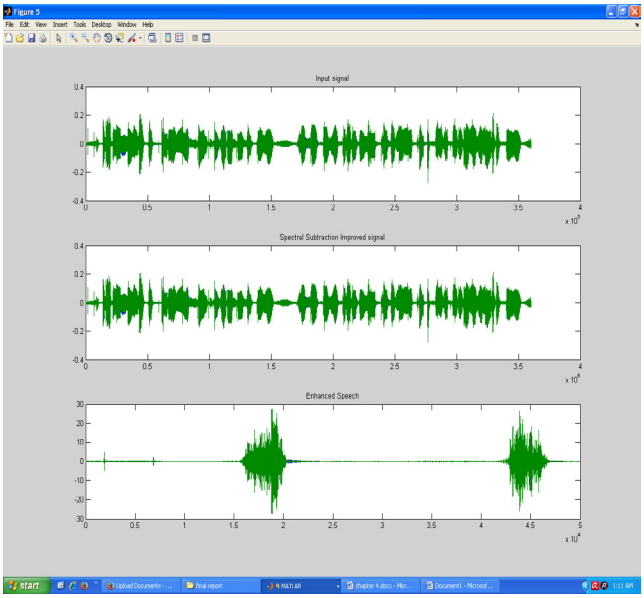

Figure 8. Spectral subtraction improved speech signal

Here Figure 8 is showing the results of implementation of spectral subtraction method over the speech generated over the signal. The Figure is showing the improved signal form obtained from this algorithmic approach.

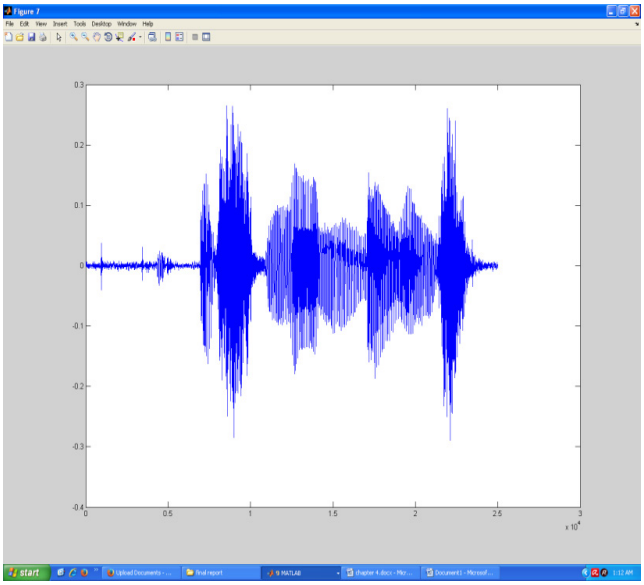

Figure 9. Processed signal form

Here Figure 9 is showing the final speech signal that is considered as the process signal for speech signal recognition.

4.2.2 Example II: Speech Signal

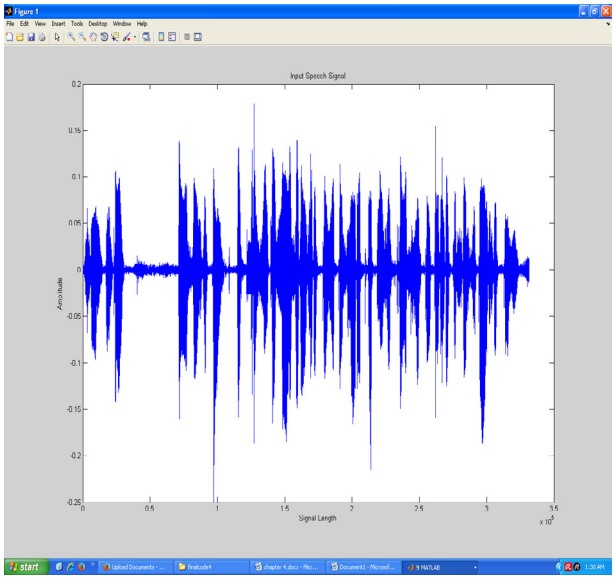

Figure 10. Raw Speech Signal

Here Figure 10 is showing the raw form of Speech signal. Here x axis represents the signal length and y axis represents the respective amplitude value.

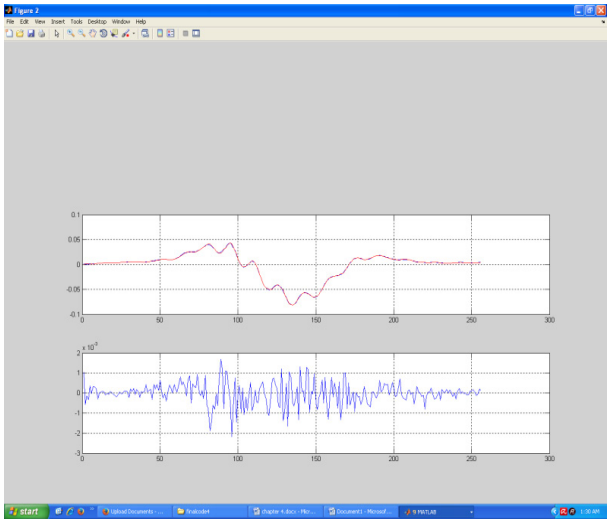

Figure 11. Result of filtration stage

Here Figure 11 is showing the results of filtration stage. The high pass filter is here applied to remove the signal noise. This stage basically removes the baseline drift noise from the signal. As the noise is removed from the signal, some delay is included in the signal itself.

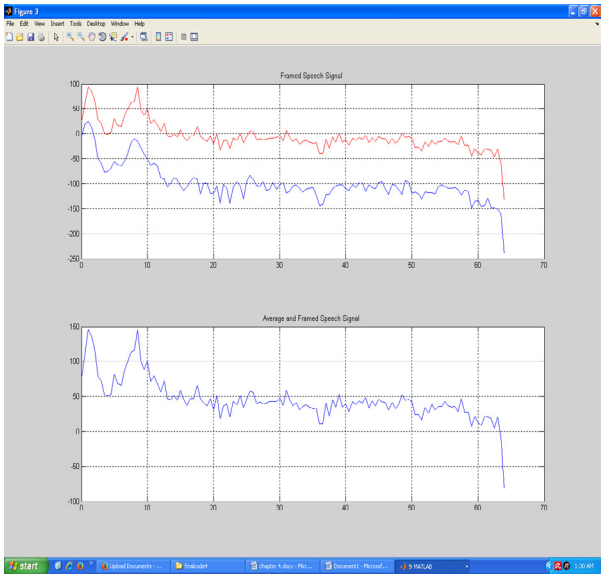

Figure 12. Frame adaptive signal analysis

Here Figure 12 is showing the mapping frame adaptive signal analysis phase so that the noise over the speech signal can be removed. The first sub plot is here defining the frame adaptation over the signal and the second is showing the mean signal value obtained after applying the frame adaptation.

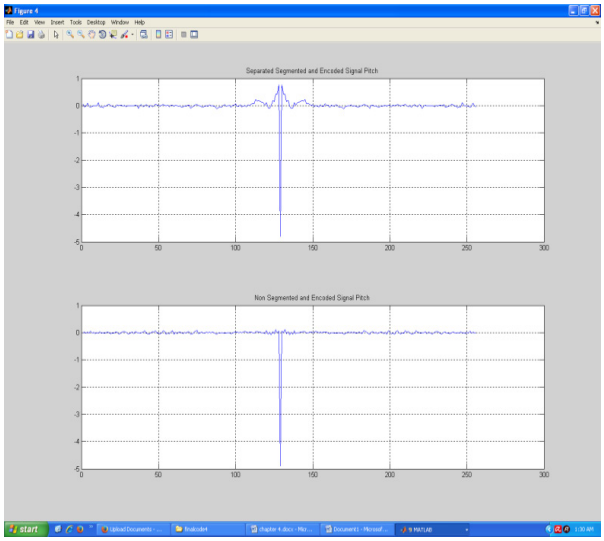

Figure 13. Separated segmented speech signal

Here Figure 13 is showing the separated segmented speech signal obtained using the segmented pitch analysis. The maximum pitch value is obtained to represent the effective signal value analysis.

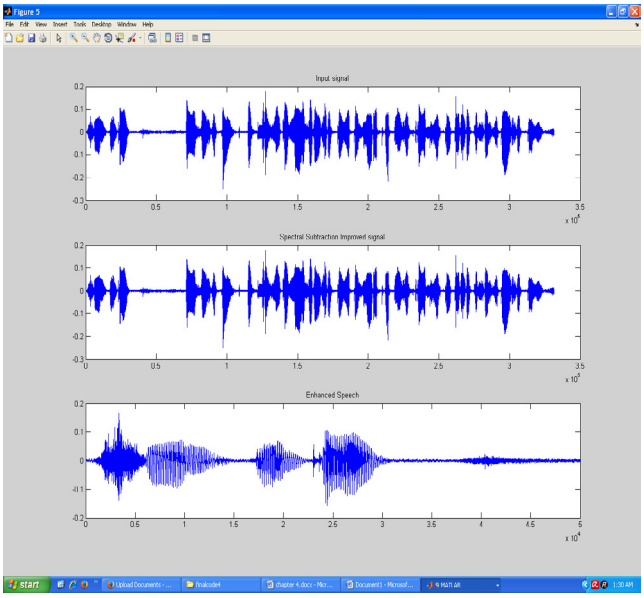

Figure 14. Spectral subtraction improved speech signal

Here Figure 14 is showing the results of implementation of spectral subtraction method over the speech generated over the signal. The Figure is showing the improved signal form obtained from this algorithmic approach.

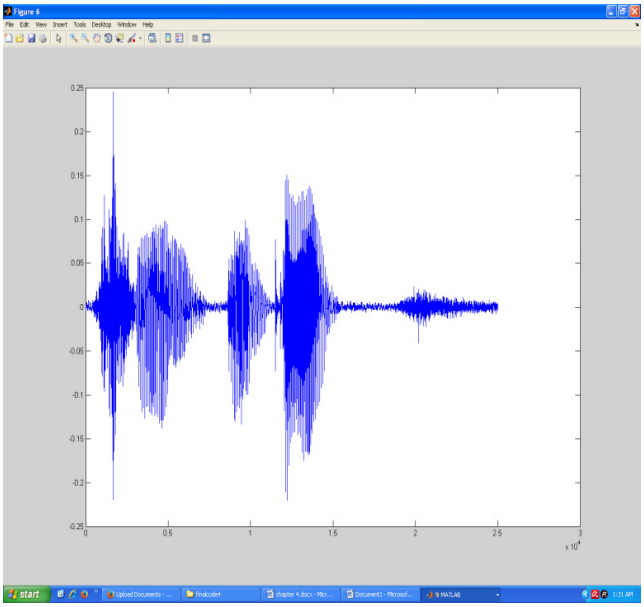

Figure 15. Processed signal form

Here Figure 15 is showing the final speech signal that is considered as the process signal for speech signal recognition.

5. Performance Analysis

The work is defined to perform the classification of speech signal classification. The properties of the dataset are given in Table 1.

Table 1. Dataset description

Properties	Values
Dataset Name	Linguistic Data Consortium for Indian Languages
Dataset URL	http://www.ldcil.org/ resourcesSpeech CorpHindi.aspx
Dataset Size	30
Format	WAV
Type	CHARACTER
Classes	CHARACTER

The work is applied on the feature set derived to take the decision about the speech signal classification. Once the feature set is generated, the dataset is divided in training and testing set. The classification is performed using SVM classifier. The parameters of this classifier are given in Table 2.

Table 2. SVM parameters

Parameter	Values
Kernel Function	Linear_kernel
Bias	0.7664
ScaleData	1x1
Group Size	32
Number of Class	8

The analysis is performed on different training and testing sets. The properties of the conducted experimentation are given below.

5.1 Dataset I

Table 3. Dataset I: classification parameters

Properties	Values
Training Set Size	30
Testing Set	14
Type	Featureset
Recognized Correctly	13
Incorrect Recognition	1
Accuracy	92.857%

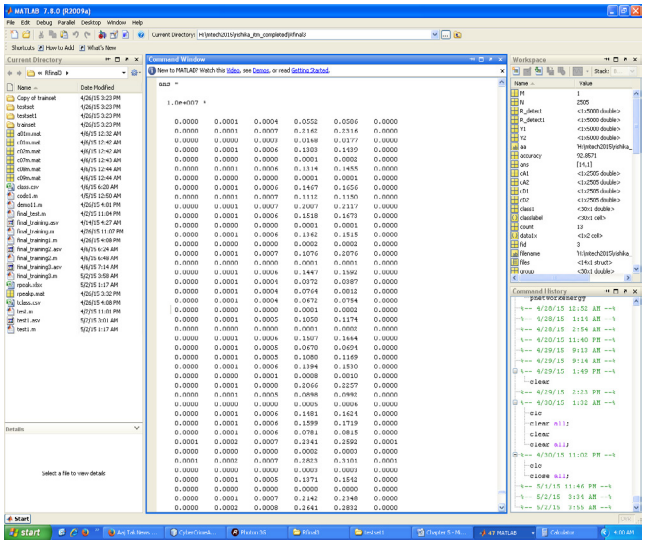

Figure 16. Feature set

Here Figure 16 is showing the feature set obtained from the work.

Table 4. Dataset I: Results (Speech Signal Identification)

Properties	Values
Total Positive	11
True Positive (Identified)	10
True Positive Rate	90.9%
False Positive	1
False Positive Rate	9.1%

Here Table 4 is showing the recognition rate obtained specifically for speech signal classification. The results show that only 1 instance is identified as wrong instance.

Table 5. Dataset I: Results (False Identification)

Properties	Values
Total Positive (False)	3
True Positive (Identified)	3
True Positive Rate	100%
False Positive	0
False Positive Rate	0%

Here Table 5 is showing the recognition rate obtained specifically for Normal patient. The results show that all the instances are identified correctly. The results are shown in the form of bar graph.

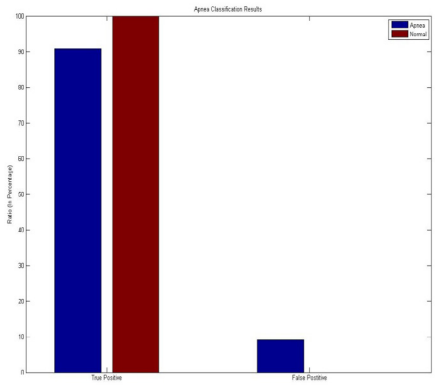

Figure 17. Analysis result

Here Figure 17 is showing the analysis results obtained from the work. The results show that the high recognition rate is obtained for valid and invalid cases.

6. Conclusion and Future Scope

6.1 Conclusion

In this work, a three layered model is presented to improve the speech recognition for Hindi characters. The presented model has used the predictive and intelligent analysis to perform the recognition. In first stage of this model, the noise reduction is performed to improve the signal strength. To improve the signal performance, DWT and spectral subtraction method are applied. These methods basically remove the background noise and other additive noise over the signal. Once the noise improved signal is obtained, the next work is to extract the signal features. The signal feature adaptation is performed using HMM based method. This work includes several statistical features including mean, standard deviation, MSE are considered. Once the feature set is obtained, the SVM is applied to perform the recognition. The experimentation of work is done on Hindi characters and words database. The results show that the work has provided effective speech signal recognition.

6.2 Future work

In this work, a feature adaptive SVM is defined to perform speech signal identification. The work can be improved in future under following aspects.

- In future, some more parameters can be taken to improve the size of feature set.
- In future, some hybrid model can be defined as classifier to improve the recognition rate.

7. References

1. Peinado AM. Use of multiple vector quantisation for semicontinuous-HMM speech recognition. *IEEE Proc-Vis. Image Signal Process*, 1994.
2. Tatsuhiko Kinjo. On Hmm Speech Recognition Based on Complex Speech Analysis. *IEEE Industrial Electronics, IECON 2006 - 32nd Annual Conference,* 6-10 Nov. 2006.
3. Cong-Thanh Do. On the Recognition of Cochlear Implant-Like Spectrally Reduced Speech With MFCC and HMM-Based ASR. *IEEE Transactions On Audio, Speech, And Language Processing.* 2010 July; 18(5):1065-68.
4. Panikos Heracleous. Analysis and Recognition of NAM Speech Using HMM Distances and Visual Information. *IEEE Transactions on Audio, Speech, and Language Processing.* 2010 Aug; 18(6): 1528-38.
5. Revathi A. Speaker Independent Continuous Speech and Isolated Digit Recognition using VQ and HMM. *Communications and Signal Processing (ICCSP), 2011 International Conference.* 10-12 Feb. 2011.
6. Jinyu Li. Improving Wideband Speech Recognition Using Mixed-Bandwidth Training Data in CD-DNN-HMM. SLT 2012, *IEEE Workshop on Spoken Language Technology*, Inproceedings, January 1, 2012.
7. Ashok Shigli. A Spectral Feature Process for Speech Recognition Using HMM With MFCC Approach. *2012 National Conference on Computing and Communication Systems (NCCCS).* 21-22 Nov. 2012.
8. Mohit Dua. Punjabi Automatic Speech Recognition using HTK. *IJCSI International Journal of Computer Science Issues.* 2012 Jul; 9(4):359-64.
9. Preeti Saini. Hindi Automatic Speech Recognition Using HTK. *International Journal of Engineering Trends and Technology (IJETT)*. 2013 Jun; 4(6):2223-29.
10. Sunija AP, Rajisha TM, Riyas KS. Comparative Study of Different Classifiers for Malayalam Dialect Recognition System. *Procedia Technol.* 2016; 24:1080-88.
11. Rajisha TM, Sunija AP, Riyas KS. Performance Analysis of Malayalam Language Speech Emotion Recognition System Using ANN/SVM. *Procedia Technol.* 2016; 24:1097-1104.
12. Hoesen D, Satriawan CH, Lestari DP, Khodra ML. Towards Robust Indonesian Speech Recognition with Spontaneous-Speech Adapted Acoustic Models. *Procedia Comput. Sci.* 2016 May; 81:167-73.