Dnyanopasak Research Journal

ISSN: (Print)

Spoken Marathi Vowel recognition using ensemble feature extraction technique

Agnihotri P.P.*, Khanale P.B.

School of Technology, S.R.T.M. University Nanded, Sub-Centre, Latur
**Corresponding authors: Agnihotri_PP@rediffmail.com, Prakashkhanale@gmail.com*

Abstract

Speech recognition is a process of converting spoken words into text. The Marathi language has 13 vowels and 36 consonants. This language is syntactically and semantically different than the global language like English. The speech recognition systems are working well in the acoustically protected environment but performance of system is degraded in the noisy environment. So it is necessary to search a more efficient feature extraction technique which gives better performance under noisy environment. This paper represents the performance of conventional and ensemble feature extraction technique through vector quantization classifier. The proposed technique is applied on spoken Marathi vowels recorded with three different speakers. It is observed that the ensemble feature extraction gives better performance than conventional techniques.

Keywords: MFCC, LPC, PLP, RASTA, Ensemble Feature extraction and Vector Quantization.

INTRODUCTION

Speech is most common and effective medium of communication for information exchange between human beings. People are using computer system for various purposes like controlling applications through commands, data entry, document preparation, data analysis, information retrieval and entertainment etc. The

recognition of speech through computer is a challenging task due to variations in different parameters of speech such as source variation, acoustic variation, physical and emotional state of speaker variation, speaking rate variation and variation in socio-linguistic background of a speaker.

Despite that speech recognition systems are working well in acoustically protected environment. The performance of SR system is degraded in noisy environment. So there is need to search a effective technique to extract the features from the speech signal which gives better performance under noisy conditions.

Feauture Extraction

Feature extraction is a process of transforming input signal into set of features that gives important information to perform the desired tasks [1]. The task shall be performed by the feature extraction are as follows

a. To analyse the speech
b. To code the speech
c. To enhance the speech
d. To recognize the speech
e. To recognize the speaker
f. To identify the language

This study focuses on Mel frequency cepstral coefficient (MFCC), Linear predictive coding (LPC), Perceptual linear prediction (PLP), Relative Spectra (RASTA) processing methods in feature extraction and combination of two or more than two techniques.

1. Mel Frequency Cepstral Coefficients (MFCC):

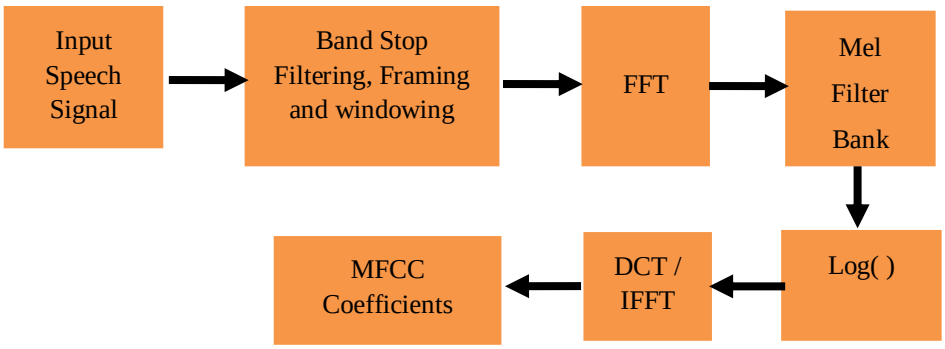


Figure 1: Flowchart for MFCC computation

It is used to identify the components of the audio signal that holds linguistic content. In human speech production system, sounds generated by a human are filtered by the shape of vocal tract including tongue, teeth etc. The shape of the vocal tract is used for determination of envelope of the short time power spectrum.The MFCC is a mean to represent such kind of envelope accurately [2]. MFCC is a feature widely used in automatic speech and speaker recognition.

Output of MFCC Algorithm

The MFCC features are extracted by following the steps as shown in the figure1which is applied on spoken Marathi vowels. The output of MFCC algorithm generates the Mel spectrum as shown in the figure 2.

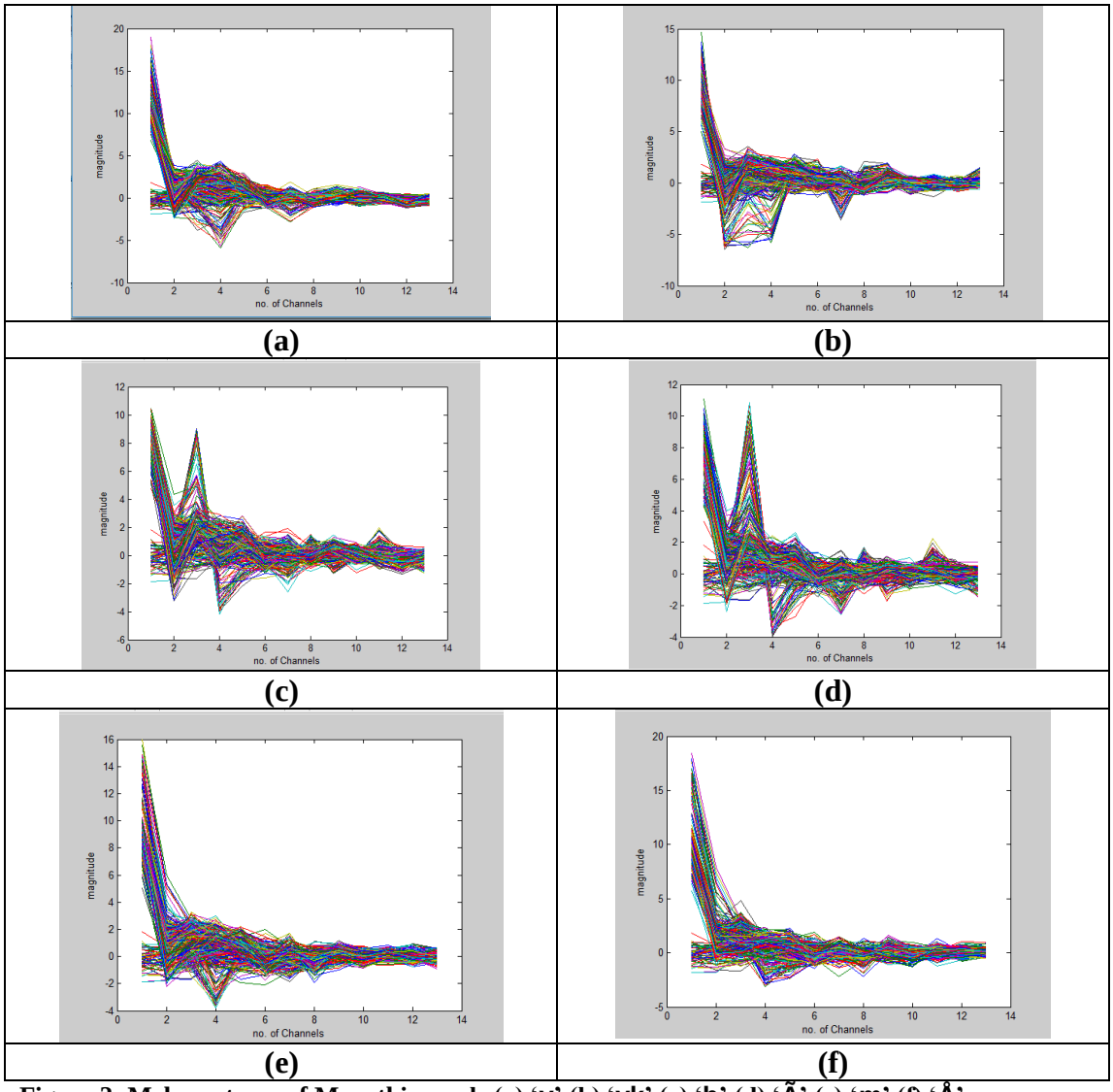


Figure 2: Mel spectrum of Marathi vowels (a) 'v' (b) 'vk' (c) 'b' (d) 'Ã' (e) 'm' (f) 'Å'

2. Linear Predictive Coding (LPC): This method is used to estimate basic speech parameters like pitch, formants, spectra and vocal tract area function. It is also used to represent the low bit rate transmission. This method is used widely because of the following reasons[5]

a) All pole model of LPC gives better prediction of vocal tract spectral envelope
b) It helps to represent the parsimonious representation of the characteristics of vocal tract
c) This method is the easiest method for implementation using software and hardware
d) The LPC model's performance for recognition of speech is better than the filter bank front ends.

LPC implementation:

The LPC coefficients are determined by following the LPC computation procedure. It includes the various phases as shown in the figure [3].

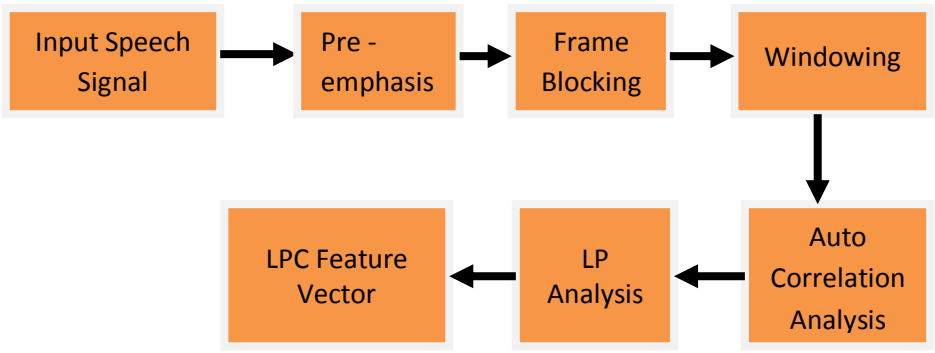


Figure 3: Flowchart for LPC Computation

Output of LPC Algorithm:

LPC has been applied on Marathi vowels and feature vector is created for each sample. The output LPC algorithm generates LPC spectrum as shown in the figure 4.

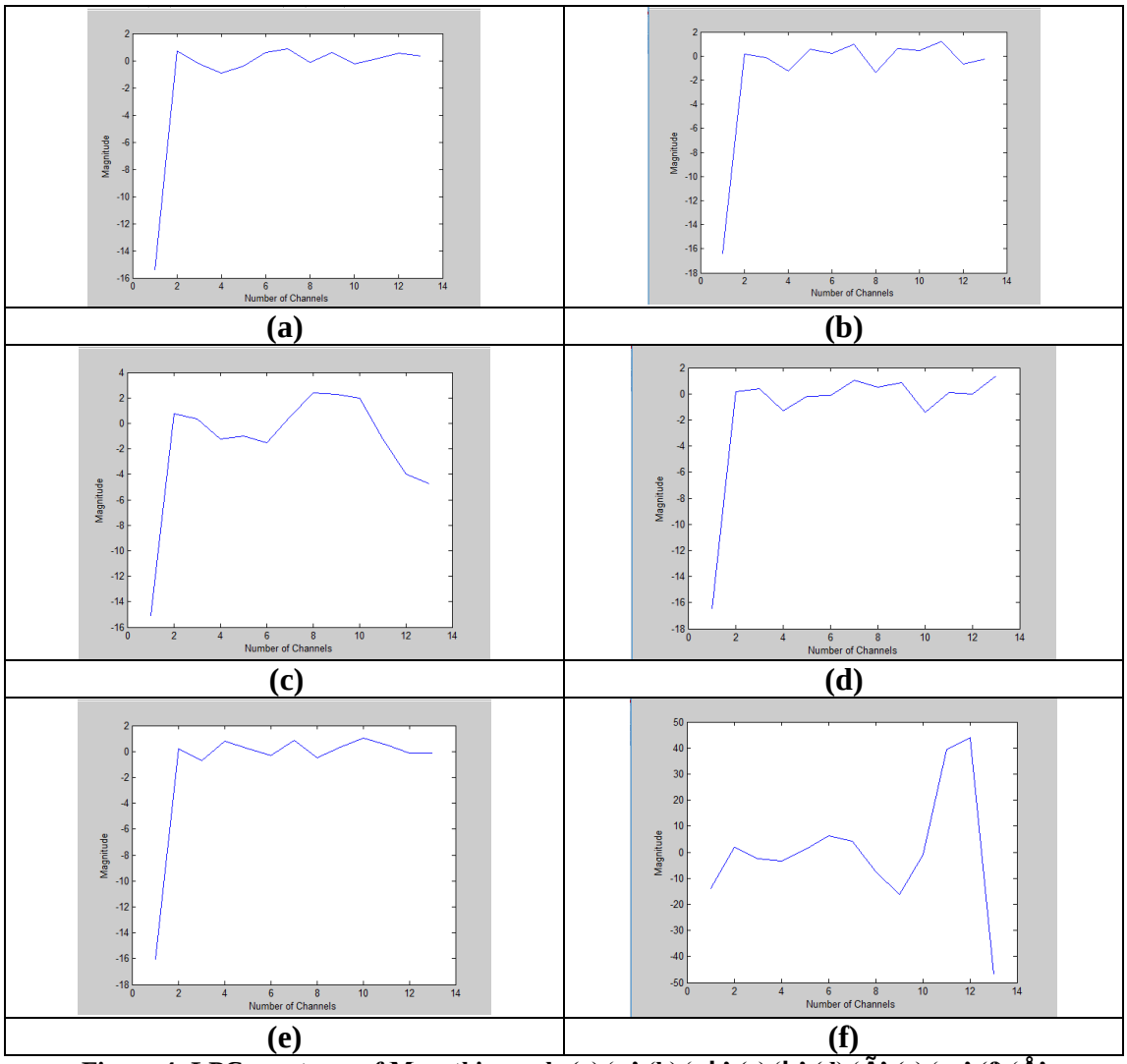


Figure 4: LPC spectrum of Marathi vowels (a) 'v' (b) 'vk' (c) 'b' (d) 'Ã' (e) 'm' (f) 'Å'

Perceptual Linear Prediction (PLP) :

In PLP models the speech is based on psychophysics hearing. PLP and LPC are similar except that its spectral characteristics have been transformed to match human auditory system [4].

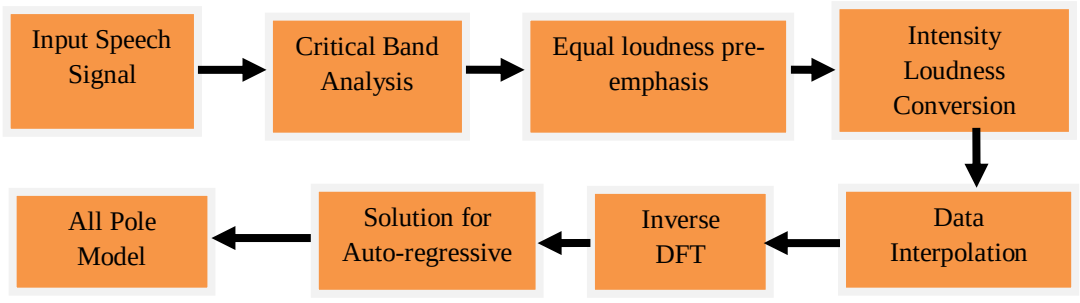


Figure 5: flowchart for PLP computation [9]

The PLP features are extracted by following the procedure as shown in the figure 5.

Output of PLP Algorithm

PLP Algorithm is applied on Marathi vowels. The output of PLP algorithm generates the power spectrum as shown in the fig. 6.

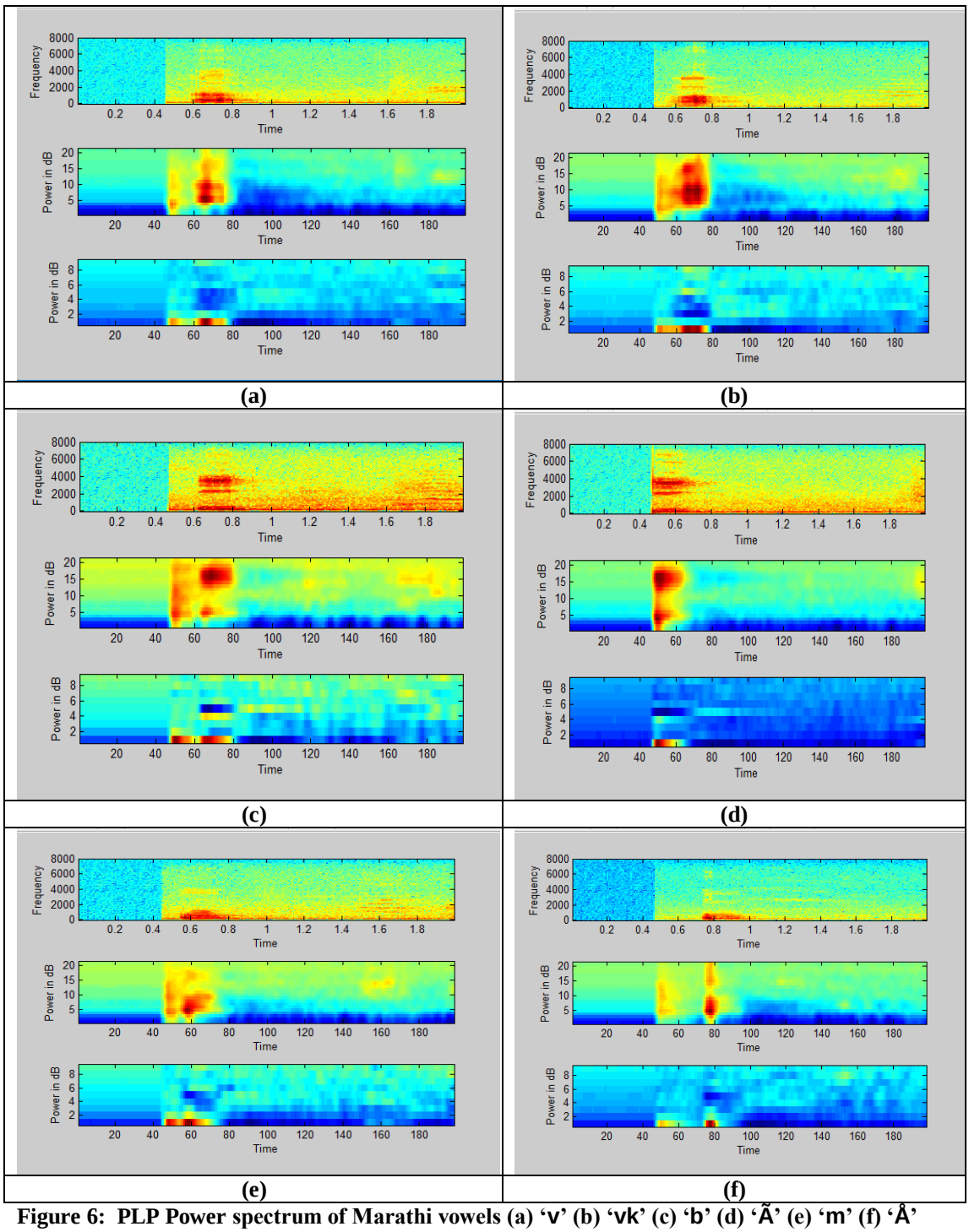


Figure 6: PLP Power spectrum of Marathi vowels (a) 'v' (b) 'vk' (c) 'b' (d) 'Ã' (e) 'm' (f) 'Å'

4. Relative Spectral Amplitude or Relative Spectra (RASTA):

RASTA Technique is proposed by Hemensky and Morgan. This technique is used to make speech patterns less sensitive to the steady state factors exists in the speech patterns while adopting any change. The RASTA technique is the extension of PLP techniques. RASTA is a designed to reduce or suppress the convolutional or additive noise. RASTA features are determined by following the procedure as shown in the figure 7.

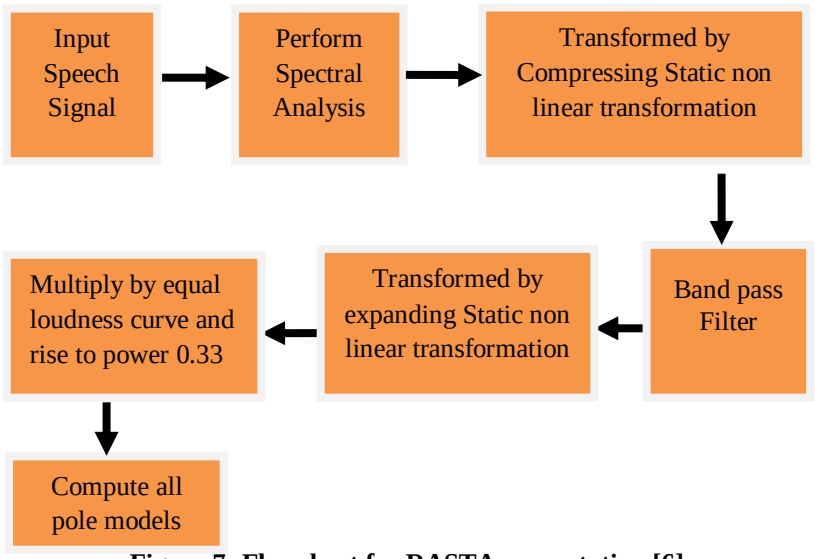
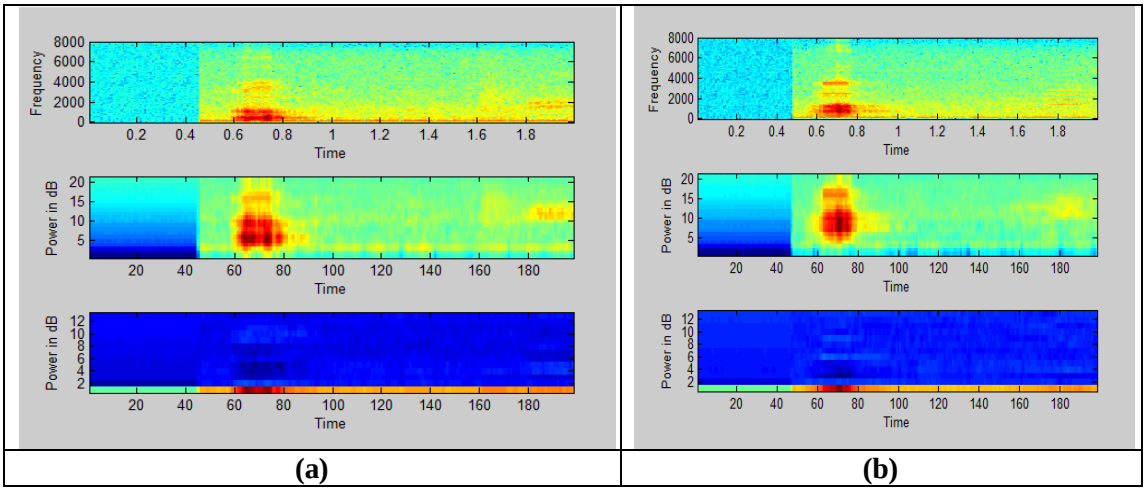


Figure 7: Flowchart for RASTA computation [6]

Output of RASTA Algorithm:

RASTA: The output of algorithm generates a power spectrum is as shown in the figure 8.



(a) **(b)**

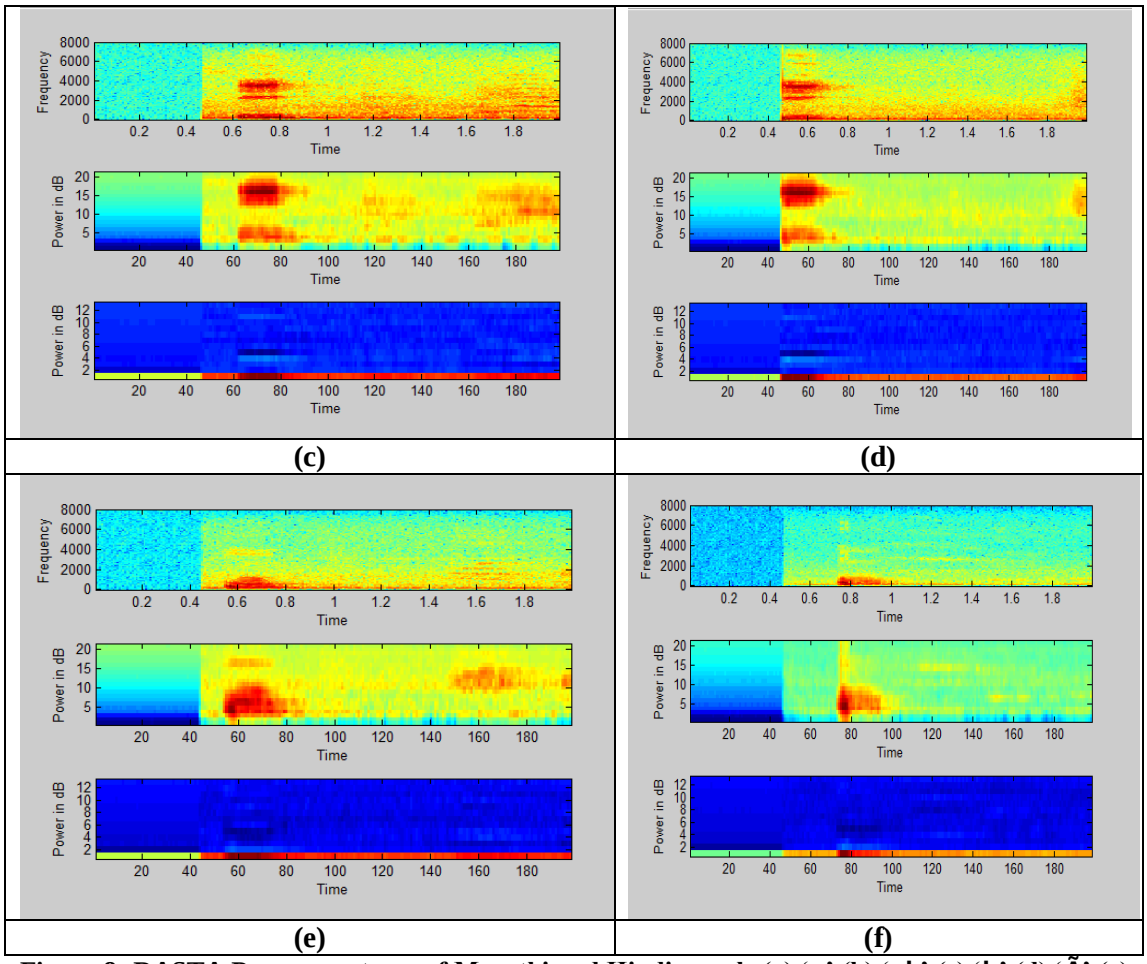


Figure 8: RASTA Power spectrum of Marathi and Hindi vowels (a) 'v' (b) 'vk' (c) 'b' (d) 'Ã' (e) 'm' (f) 'Å'

5. Ensemble feature extraction technique:

First of all the features of speech samples are determined using MFCC, LPC, PLP and RASTA methods separately. It is found that some patterns are recognized by one method but these patterns are not recognized by the other methods. For example the pattern of '**V**' is not recognized by using PLP features but it is recognised by using MFCC features. So there is need to combine the two methods to improve the accuracy of recognition of speech samples.

The two or more than two methods are ensemble on the basis of feature level combination of methods.

Feature classification using Vector Quantization:

It is a procedure to encode the input vector into the whole numbers. The collection of input vectors into matrix is called as codebook. In this codebook, each

row represents the feature of a single sample. The number of rows of the codebook represents the number of samples used for training a vector quantizer. The spoken of Marathi vowels were prepared as a training set of vector quantizer.

This method is used to classify the patterns. This process is divided into two parts: Feature training and Feature matching. In feature training process, it trains the system using the codebook which have input feature vectors. The feature matching process is done by measuring the similarity between feature extracted from unknown patterns and feature vectors of codebook [7].

The distance between unknown pattern and the reference pattern can be calculated using Euclidian distance measure technique.

The algorithm is designed for recognizing Marathi spoken vowelss using ensemble feature extraction techniques. The proposed algorithm is as follows:

PROPOSED ALGORITHM:

Inputs Feature vectors of MFCC,LPC,PLP and RASTA

Output Recognition accuracy of speech corpora

Step 1 Prepare the feature vector codebook of two or more than two feature vectors of standard database as well as test database like

Ystd←[M1 L1] or [P1 R1] or [M1 P1 R1]

Ytest←[M1 L1] or [P1 R1] or [M1 P1 R1]

Step 2 Determine the Euclidian distance between Ystd and Ytest

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (Ystd_{ik} - Ytest_{jk})^2}$$

Step 3 Set the threshold value for the recognition accuracy

T ← 0.2

Step 4 Prepare the resultant matrix using threshold value T

Repeat the step while(i<=n)

Repeat the step while(j<=n)

If(d(i,j)<=T) then Set res(i,j)=1

Otherwise Set res(i,j)=0

(end of step 4)

Step 5 Determine the number of samples correctly recognized

Set acc=0

Repeat the step while(i<=n)

Repeat the step while(j<=n)

If(i=j) then

If(res(i,j)=1) then

acc=acc+1

(end of step 5)

Setp 6 Determine the Percentage of accuracy

Accuracy=acc/N

(Where N is total number of samples)

Results and Discussion:

All feature extraction techniques were implemented in MATLAB 2013a. First of all the feature are extracted for four different methods individually and there after feature level combination of two or more than two methods is made to improve the accuracy. The comparison of individual and ensemble feature extraction is done on the basis of correct classification of the Spoken Marathi Vowels. Table 1 shows the comparison of various feature extraction technique on the basis of recognition accuracy.

Table 1: Comparison of feature extraction techniques on the basis of recognition accuracy

Sr. No	Feature extraction method	Number of samples	No. of correctly classified samples	Accuracy of recognition (in %)
1.	MFCC	39	35	89.74
2.	LPC	39	33	84.61
3.	PLP	39	31	79.48
4.	RASTA	39	34	87.17
5.	MFCC+LPC	39	35	89.74
6.	MFCC+PLP	39	36	92.30
7.	MFCC +RASTA	39	37	94.87
8.	MFCC+PLP+RASTA	39	38	97.43
9.	LPC+PLP	39	35	89.74
10	LPC+RASTA	39	36	92.30
11	LPC+PLP+RASTA	39	35	89.74

CONCLUSIONS

It is found that MFCC has highest recognition accuracy as an individual method i.e. 89.74 % where as LPC has 84.61%, PLP has 79.48% and RASTA has 87.17%

It is observed that a combination of MFCC, PLP and RASTA gives highest recognition accuracy of spoken Marathi vowels i.e. 97.43%.

It is concluded that the feature level combination of feature extraction methods gives better results as compared to conventional methods.

REFERENCES

1. Veton Z. Këpuska, Hussien A. Elharati, 2015, "Robust Speech Recognition System Using Conventional and Hybrid Features of MFCC, LPCC, PLP, RASTA-PLP and Hidden Markov Model Classifier in Noisy Conditions" , Journal of Computer and Communications, vol. 3, pp 1-9.
2. Vimala.C, Radha.V, 2014 "Suitable Feature Extraction and Speech Recognition Technique for Isolated Tamil Spoken Words"International Journal of Computer Science and Information Technologies, Vol. 5(1), 2014, 378-383 ISSN No. : 0975-9646.
3. Pawan Kumar, Astik Biswas, A.N. Mishra and Mahesh Chandra,2010 "Spoken Language Identification using Hybrid Feature Extraction Methods" , Journal of Telecommunications, Vol. 1, Issue 2.
4. Sai Bharadwaj B , U.Vagdevi and Yadav Bineetha, 2013 Linear Prediction Analysis, The International Journal Of Engineering And Science (IJES), Volume2 Issue 4 pp01-07 ISSN(e): 2319 – 1813
5. Namrata Dave,2013, "Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition", International Journal For Advance Research In Engineering And Technology, Volume 1, Issue 6, pp 1-5
6. Kashyap Patel, R.K. Prasad, 2013, "Speech Recognition and Verification Using MFCC & VQ", International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319– 6378, Volume-1, Issue-7, pp 33-37
7. Dipmoy Gupta, Radha Mounima C. Navya Manjunath, Manoj PB, 2012, "Isolated Word Speech Recognition Using Vector Quantization (VQ)", International Journal of Advanced Research in Computer Science