Comparison of Different Speech Feature Extraction Techniques with and without Wavelet Transform to Kannada Speech Recognition

M.A.Anusuya
Sr.Lecturer
Dept.of CS&E,
SJCE,Mysore-06

S.K.Katti
Emeritus Professor
Dept.of CS&E,
SJCE, Mysore-06

ABSTRACT
Pre-processing of speech signals is considered a crucial step in the development of a robust and efficient speech or speaker recognition system. This paper deals with different speech processing techniques and the recognition accuracy with respect to wavelet transforms. It is shown that by applying wavelet transform to the conventional methods the signal recognition accuracy will be increased by using discrete wavelet transforms and the wavelet packets for clean and noisy speech signals respectively. Results presented in the tabular form, shows the advantage of pre-processing the signals with wavelet techniques gives good results over conventional methods.

Keywords
Speech signal, pre-processing, Discrete Wavelets Transforms(DWT), Wavelet packet decomposition (WPD),Linear Predictive co-efficient (LPC), kannada, isolated words, Mel frequency cepstral co-efficient (MFCC), RelAtive Spectral Transform- Perceptual Linear Prediction approach (RASTA-PLP) ,Euclidean distance**.**

1. INTRODUCTION
Speech is a unique form of audio data. It is a relatively simple and widely studied type of acoustic signal. Pre-Processing of speech signals, i.e. segregating the voiced region from the silence/unvoiced portion of the captured signal is usually advocated as a crucial step in the development of a reliable speech or speaker recognition system. This is because most of the speech or speaker specific attributes are presented in the voiced part of the speech signals [1]; moreover, extraction of the voiced part of the speech signal by marking and/or removing the silence and unvoiced region leads to substantial reduction in computational complexity and increases the recognition accuracy of the speech signal at later stages of feature extraction [2,1].

One of the accepted ways of labeling a speech signal is the three state representation: (i) Silence region (S) where no speech is produced, (ii) Unvoiced region (U), where the resulting waveform is a periodic or random in nature as the vocal chords do not vibrate, and (iii) Voiced region (V) where the resulting waveform is quasi-periodic as the vocal chords are tensed and hence vibrate periodically [3, 1]. It should be made clear that the segmentation of the speech signal in the aforementioned regions is not very rigid; however, it has been noted that small errors in the boundary locations seldom have any significant effect in most of the applications [3].

In this paper all the conventional methods of the feature extraction techniques has been discussed and the preprocessing of the speech signal is done with wavelets. For the clean speech DWT is used for pre-processing the signal and on these, conventional methods have been applied and the results are observed. For noisy speech, the WPD is applied for pre-processing and noise removal, and on his the conventional methods has been applied, and the results are observed. The paper has been organized as follows: Section II talks about the wavelet Transform technqiues i.e. Discrete wavelet transforms and the wavelet packet transforms. Section III shows the data base construction procedure, Section IV shows the algorithmic procedures using LPC and MFCC, RASTA-PLP methods with wavelet transforms. The obtained results and graphs are discussed in section V, and finally conclusions are drawn in section VI.

2.WAVELET TRANSFORM TECHNIQUES:

2.1. THE DISCRETE WAVELET TRANSFORM:

The DWT can be used for Multi Resolution Analysis (MRA)[4,5]. Since speech signal is a non-stationary and non-linear signal MRA can be used. The given signal is decomposed into the approximation and detail coefficients. A given function f(t) satisfying certain conditions[4], can be expressed through the following representation[6].

$$f(t)= \sum_{j=1}^{L} \sum_{k=-\infty}^{\infty} d(j,k)\varphi(2^{-j}t-K) + \sum_{k=-\infty}^{\infty} a(L,K)\theta(2^{-L}t-K)$$

Where $\psi(t)$ is the mother wavelet and $\theta(t)$ is the scaling function. a(L,k) is called the approximation coefficient at scale

L and d(j,K) is called the detail coefficient at scale j. The approximation and detail coefficients can be expressed as

$$a(L, K) = \frac{1}{\sqrt{2^L}} \int_{-\infty}^{\infty} f(t)\theta(2^{-L}t - K)dt$$

$$d(j, K) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} f(t)\varphi(2^{-j}t - K)dt$$

The DWT can be viewed as the process of filtering the signal using a low pass (scaling) filter and high pass (wavelet) filter [7,8]. Thus, the first layer of the DWT decomposition of a signal splits it into two bands giving a low pass version and a high pass version of the signal. The low pass signal gives the approximate representation of the signal while the high pass filtered signal gives the details or high frequency variations. The second level of decomposition is performed on the low pass signal obtained from the first level of decomposition (as shown in Fig.1). The wavelet decomposition of the signal *S* analyzed at level *j* has the following structure: [*cAj, cDj, . . . , cD*j].For more information about the Wavelet Transform (WT) refer [11,12]. Based on the choice of the mother wavelet ψ(t) and scaling function θ(t), different families of wavelets can be constructed[4,5,9,10]. Daubechies wavelets with different decomposition levels namely: Db8 –level5 and Db10-level5 can been used. In this paper, only DB8-levl5 is considered for the experimental purpose, since it is best among the Daubecies family[19,20].

2.2 The Wavelet Packet Decomposition

The *wavelet packet* method is a generalization of wavelet decomposition that offers a richer range of possibilities for signal analysis. Wavelet packet atoms are waveforms indexed by three naturally interpreted parameters, position, scale, frequency. In wavelet packet analysis each detail coefficient vector is also decomposed in to two parts using the same approach as in approximation vector splitting. The information lost between two successive approximations is captured in the detail coefficients. Then the next step consists of splitting the new approximation coefficient vector; successive details are never reanalyzed. In the wavelet packet situation, each detail coefficient vector is also decomposed into two parts using the same approach as in approximation vector splitting. This yields more than different ways to encode the signal. This offers the richest analysis [13]. The complete binary tree is produced. The wavelet packet decomposition is shown in the figure 1. In the WPD, both the detail (cHj, cVj, cDj) and approximation coefficients are decomposed. WPD coefficients are used for dual purpose ie. for feature coefficients and as a technique for removing the noise from noisy speech.

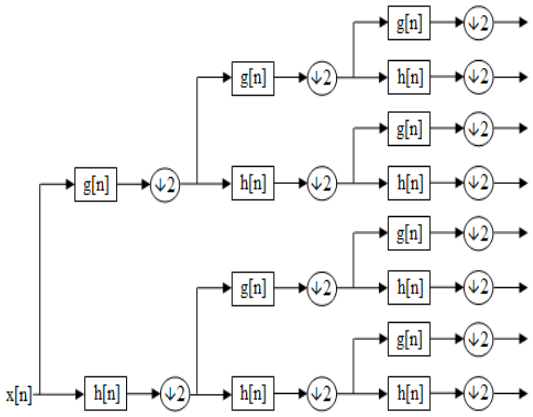

Figure 1 wavelet packet decomposition

3. SPEECH DATABASE CONSTRUCTION AND WPD COEFFICIENTS COMPUTATION

Table 1

Number	Kannada Word	Symbol used in the paper
1	"Ondu"	One
2	"Eradu"	Two
3	"Muru"	Three
4	"Nalaku"	Four
5	"Iydu"	Five
6	"Aaru"	Six
7	"Yelu"	Seven
8	"Enttu"	Eight
9	"Ombatthu"	Nine
10	"Hatthu"	Ten

Database contains 10 female speakers. Each speaker utters each kannada digit 10 times. The leading and trailing silence is removed from each utterance. All samples are stored in Microsoft wave format files with 8000 Hz sampling rate, 16 bit PCM and mono channels. In our experiments, training and testing is performed on clean and noisy speech utterances. To acquire the speech signal, PRAAT software is used. 10 adult female speakers were asked to utter the 10 kannada words individually and separately. Totally, 1000 signals are collected from all the speakers.

The signals are considered as noisy signals as they have acquired in the normal noisy environment. i.e use of fan, phone, etc. i.e. room noise is considered. General noise itself is considered as noisy signal.

3.1Construction of Database for Training purpose:

An adult female, native speaker of Kannada was asked to utter the Kannada words(1 through 10, see table-1), and her voice

was sampled at 8KHz. The speech signal of each word was then isolated from silence. The signal is decomposed using DWT and WPD coefficients for clean and noisy speech signal. The clean speech signal samples are decomposed upto level 5 using DWT and then stored in ascending order: Firstly, the ten samples corresponding to word one("Ondu") were stored, then the ten samples of two and so on. For noisy signal, the samples are decomposed up to 5^{th} level using WPD ,and the signal samples are stored in ascending order as done in clean speech. These stored signals are used in training phase of the experiment.

3.2. Construction of Database for testing purpose:

Separate 500 samples are collected for both clean and noisy speech signals for testing purpose from 10 female speakers. Each word is uttered 5 times, from all the speakers, totally 500 signals are used for testing purpose.

Each of the 1000 speech samples are decomposed into DWT/WPD coefficients depending on the input of the speech signal. Both DWT and WPD is calculated up to 5 levels. For n=5 levels of decomposition the WPD produces 2^n different sets of coefficients. Shannon entropy is used for Wavelet Packet Decomposition. In each level, DWT/WPD decomposes the signals without losing the integrity of the signal by splitting it into its approximation coefficient and detailed coefficient. Decomposition is carried out on each of the 1000 speech samples, using Daubechies wavelet i.e. Db8-Levl5[14,15,19,20] as they have been reported to be highly successful in speech compression schemes using wavelets[16] . WPD coefficients are calculated upto 5^{th} level using Shannon entropy in wavelet packet decomposition. Figure 2 shows the signal decompositions applied to different feature extraction techniques.

- Daubechies Wavelets[9]
  Daubechies8, 5-Level decomposition (db8,Lev5)

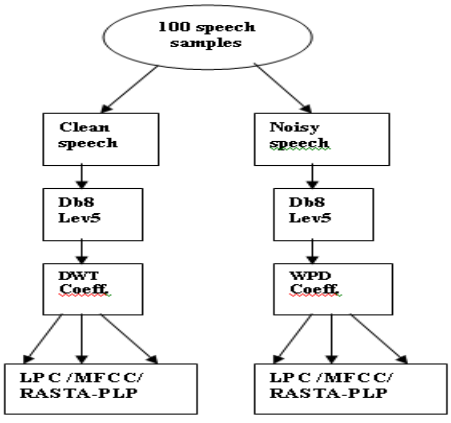

Fig.2 Decomposition of speech signal using DWT/ WPDs

4. Calculation of wavelet coefficients using (DWT/WPD) Discrete Wavelets and Wavelet packets for LPC, MFCC and RASTA-PLP methods:

The LPC/MFCC coefficients from the DWT/ WPD coefficients, are calculated using conventional methods. Before applying the conventional methods the signal has been decomposed into DWT/WPD coefficients.

4.1. Computations of LPCC Co-efficients

The DWT/WPD coefficients of each speech signal are arranged in descending order, starting from the corresponding highest level approximation coefficients followed the same levels detail coefficients followed subsequently by lower levels detail coefficients in descending order.

The fist step is pre-emphasis which basically makes one sample in speech influence the next sample by a certain weight.

$$S_1(n) = s(n) - a*s(n-1)$$

- Then, the DWT/WPD coefficients are framed into frames of 160 samples in length
- Overlap between successive frames is kept at 80 samples
- Each frame is multiplied by a 160 point Hamming window. This step is primarily to have a smooth transition between samples of a frame.
- LPC coefficients of the order 9 has been calculated

4.2. Computations of MFCC Co-efficients:

To Compute MFCC coefficients from the WPD coefficients, we employ the following method.

- The DWT/WPD coefficients of each speech signal are arranged in descending order, starting from the corresponding highest level approximation coefficients followed the same levels detail coefficients followed subsequently by lower levels detail coefficients in descending order.

- The fist step is pre-emphasis which basically makes one sample in speech influence the next sample by a certain weight.

$$S_1(n) = s(n) - a*s(n-1)$$

- Then, the WPD coefficients are framed into frames of 160 samples in length
- Overlap between successive frames is kept at 80 samples.
- Each frame is multiplied by a 160 point Hamming window. This step is primarily to have a smooth transition between samples of a frame.
- FFT to obtain the magnitude frequency response of each frame.
- The next step is to pass the data through Mel filters. Mel filters are triangular equidistant filters in Mel scale, which is a logarithmic scale.

- Taking logarithm of this we obtain Mel spectrum Co-efficients.

- The final step in obtaining MFCC is performing discrete cosine transform (IDCT) on the Mel-spectrum coefficients. The output of IDCT is Mel-cepstral coefficients of 13^{th} order.

Usually first twelve of these coefficients are used along with the energy coefficient to form the MFCC vector of length 13.

4.3. Computations of RASTA-PLP Co-efficient:
The same procedure has been applied to Critical band computation of the noisy speech signal with the same input coefficients taken from WPD coefficients for noisy speech signals, and the parameters are calculated using RASTA-PLP method. For the noisy signal, first the signal is de-noised and decomposed using WPD. For these obtained features the RASTA-PLP technique is applied. The importance of using PLP is that it tries [16], to model a perceptually motivated spectrum by an all pole model function using the autocorrelation LP technique. The RASTA (RelAtive SpecTrA) approach [17,18] is based on a band-pass time-filtering applied to a log-spectral representation of the speech, such as the log filter bank energies. Our aim is to evaluate the effect of the Rasta filtering coefficients by applying wavelet technique.

5. EXPERIMENTAL RESULTS AND GRAPH ANALYISI

We obtain four sets of LPC/MFCC/RASTA-PLP coefficients. This is calculated using DWT and WPD coefficients for clean and noisy speech signals for different wavelet families. Each of these sets has 1000 entries. Each entry is actually the collection of DWT/WPD coefficients of the speech signal from which it was derived.

i)At the end of the LPC method, we get four sets of LPC coefficients each of which has 1000 rows (corresponding to each of the 10 utterances of each word) In order to get the LPC coefficients the 10^{th} order LPC coefficients are calculated. and used in this experiment.

ii) At the end of the MFCC method, we get four sets of MFCC coefficients for each type of wavelet family. For each For 13^{th} order MFCC coefficients are calculated.

iii) For RASTA-PLP coefficients, the noisy speech signal decomposed with wavelet packets are used. The 5^{th} order autocorrelation coefficients are calculated by applying the signals to the Band pass filters. This procedure is repeated for all the signals.

Once the coefficients are calculated mean of the each group of the signals are calculated. Totally 10 means for all the ten isolated words are calculated.(i.e. one through ten). A test signal is compared with the each group mean value using Euclidian distance measure. Among the 10 groups, the test signal which gives the minimum distance is declared as the identified signal. Then the test signal is categorised to the respective group. This is repeated for all the types of the wavelet families. The results are discussed by plotting the graph in section VI.

5.1 Testing isolated words:

To test a given Kannada word, we first find its DWT and WPD coefficients. DWT is performed for clean speech and WPD is performed for noisy speech. Then LPC/MFCC/RASTS-PLP coefficients of the test signal are calculated. For these coefficients Euclidian distance is applied. Testing is performed for 5 different samples of each word, taken from the test database. So, total of 50 different test samples are used for testing all the 10 words. This procedure is repeated for all types of wavelet families and the signal is recognised

6. RESULTS

A) LPC, MFCC AND RASTA-PLP WITH AND WITHOUT WAVELET FOR CLEAN SPEECH

Table 2

Kannada Word	LP	MF	RAS-PLP	W+ LPC+ db8 LEV5	W+ MF db8 LEV5	W+ RAS-PL P db8 LEV5
"Ondu" (One)	80	90	40	80	100	50
"Eradu" (Two)	80	90	50	90	100	50
"Muru" (Three)	80	80	40	80	90	50
"Nalakku "(Four)	80	90	50	90	90	60
"Iydu" (Five)	80	80	40	80	90	50
"Aaru" (Six)	70	70	60	80	90	50
"Yellu" (Seven)	70	80	60	80	90	50
"Enttu" (Eight)	80	80	50	80	100	50
"Ombath u"(Nine)	70	70	50	80	90	50
"Hatthu" (Ten)	70	80	50	80	100	60

LP-LPC Method
MF- MFCC method
RAS-PLP- RASTA-PLP Method
W+LPC - Wavlet LPC
W+MF - Wavelet MFCC
W+RAS-PLP – Wavelet RASTA-PLP

B) LPC, MFCC AND RASTA-PLP WITH AND WITHOUT WAVELET FOR NOISY SPEECH

Table 3

Kannada Word	LP	MF	RAS - PLP	WPD + LP db8 LEV 5	WPD+ MFC+ LP db8 LEV5	WPD+ RAS -PLP db8 LEV5
"Ondu"	50	60	70	60	70	80

(One)						
"Eradu" (Two)	60	60	80	70	70	90
"Muru" (Three)	60	70	70	80	70	90
"Nalakk u"(Four)	50	60	60	60	60	80
"Iydu" (Five)	50	60	70	60	70	80
"Aaru" (Six)	40	50	70	60	60	80
"Yellu" (Seven)	50	60	80	50	70	80
"Enttu" (Eight)	60	60	70	70	60	90
"Ombath u"(Nine)	50	60	80	70	70	80
"Hatthu" (Ten)	60	60	80	70	70	80

LP-LPC Method
MF- MFCC method
RAS-PLP- RASTA-PLP Method
WPD+LPC - Wavlet Packet LPC
WPD+MFCC- Wavelet Packet MFCC
WPD+RAS-PLP – Wavelet Packet RASTA-PLP

Table 2 shows the success percentage of each of the three types of feature extraction methods for clean speech with discrete wavelet decompositions. Fig.3a and 3b shows the average success percentage of each word over all the three feature extraction methods using the DWT for clean speech. Fig.3b shows less recognition accuracy for RASTA-PLP method. The recognition accuracy of words in this method is far far lesser than other methods as shown in table 2. Table 3 shows the success percentage of each of the three types of feature extraction methods for noisy speech with wavelet packet decompositions. Fig.4a and 4b shows the average success percentage of each word over all the three feature extraction methods. Fig.4b shows less recognition accuracy for LPC and MFCC method. The recognition accuracy of words in these methods is lesser than the RASTA-PLP method as shown in table 3.

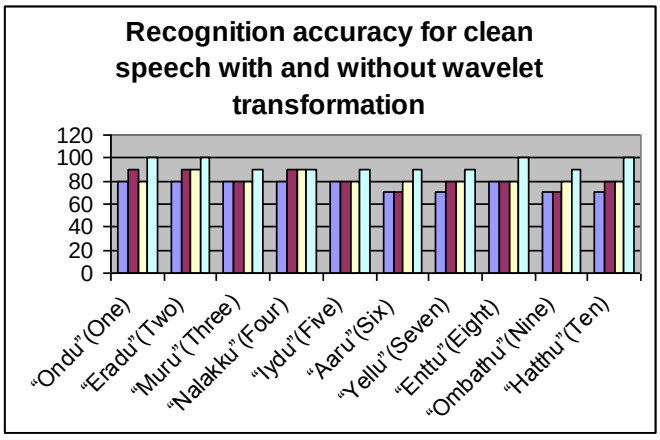

Fig.3a average success percentage of each word for the clean speech without RASTA-PLP method

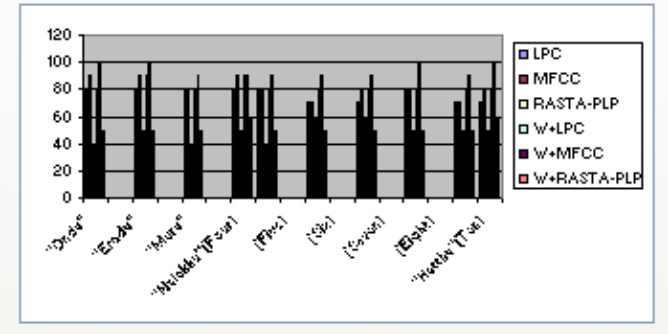

Fig.3b. average success percentage of each word for the clean speech with RASTA-PLP method

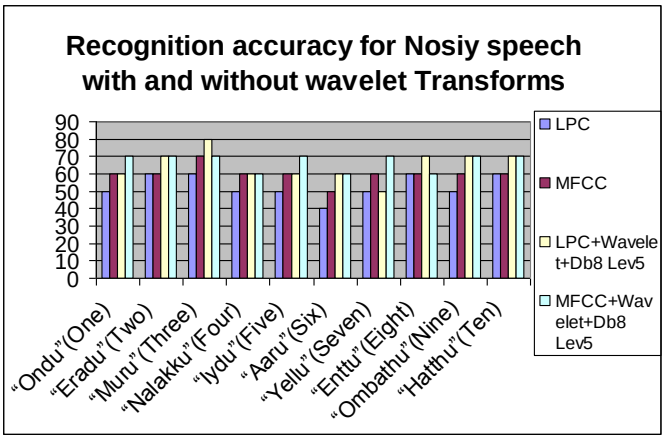

Fig 4a. average success percentage of each word for the noisy speech without RASTA-PLP method

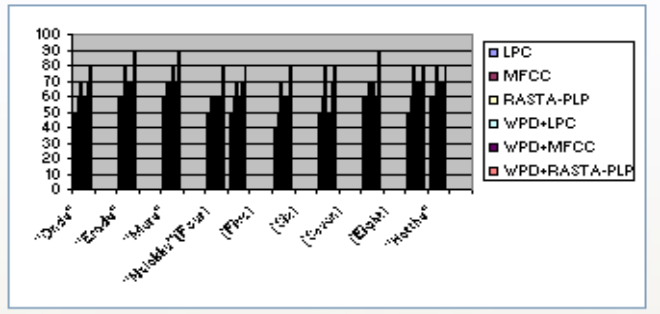

Fig 4b. Average success percentage of each word for the noisy speech with RASTA-PLP method Wavelet Packets coefficients for all the techniques

7. CONCLUSIONS

The speech signal pre-processing is carried out by different methods. Paper shows that, the feature extracted with wavelet transforms has the highest accuracy for recognition of words. This has been proved in both the conditions for clean and noisy speech signals. It is also shown that the RASTA-PLP method used with Wavelets for noisy speech yields better results. It is clearly shown form the paper, that, if the speech signals are

noisy then LPC and MFCC feature extraction methods are not the good choice. As it is seen it gives less recognition accuracy. But these methods provide good recognition results for clean speech with wavelets. If the clean speech is given then recognition accuracy for RASTA-PLP feature extraction method gives less accuracy. It is showed that, using wavelets the recognition accuracy can be increased for both clean and noisy speech signals. In particular Daubechies8, 5-level decompositions [19,20] gives the highest percentage of success in the recognition of Kannada speech. Clearly, it emerges as the candidate of choice for our DWT/WPD based speech recognition scheme, the Daubechies8, 5-level decomposition is the best wavelet family. Hence, the paper shows that, any feature extraction method with wavelet yields, good recognition accuracy of the speech signals.

8. ACKNOWLEDGEMENT

We thank Mrs.Vijayalakshmi, Asst. Professor. Department of Computer Science and Engineering, SJCE, Mysore for useful discussion with her.

9. REFERENCES:

[1] Mark Nelson ,"The Data compression Book" ,BPB publications 2,edition, ISBN 81-7029-729-X. 2003.

[2] Kalid Sayood ,"Introduction to data Compression", Morgan Kaufmann Publishers 2edition 2005.

[3] N.Venkatesh, B.Chethananand, "Tutorial on Kannada speech Recognition using Wavelet and LPC".

[4] Gilbert Strang and Truong Nguen," Wavelets and Filter Banks", Wellesley-Cambridge Press,MA,1997,pp.174-220,365-382.

[5] Andrew K.Chan and Jaideva C.Goswami, "Fundamentals of wavelets", Wiley-India, Edition, John Wiley & sons Inc, New Delhi,1999,pp.89-97.

6] Shivesh Ranjan, "A Discrete Wavelet Transform Based Approach to Hindi Speech Recognition", International Conference on Signal Acquisition and Processing, 2010.

[7] Nikhil Rao,"Speech compression using wavelets, ELEC 4801 THESIS PROGECT. School of Information Technology and Electrical Engineering, The University Of Queensland, October 2001.

[8] Brain Gamulkiewiez and Michael Weeks, " Wavelets based speech recognition", Proc. IEEE International Symposium on MicroNano Mechatronics of Human Science, Dec.2003,pp.678-681 Vol.2, doi:10.1109/ MWSCAS.2003.1562377.

[9] Ingrid Daubechies, "Ten Lecturers on Wavelets", SIAM, 1992,pp.115-132,194-292,258-259.

[10] Martin Vettereli and Jelena Kovacevic, " Wavelets and

Sub-band Coding" Prentice Hall, 1995,pp.233-238.

[11]. M. Vetterli and J. Kovacevic, "Wavelets and sub band coding", Prentice Hall, Englewood Cliffs, NJ, USA, 1995.

[12] S. Mallat, "*A wavelet tour of signal processing"*, Academic Press, 1998.

[13] Y.T.Chan "Wavelet Basics", Kulwer Academic Publications,©1995.

[14] J.S.Walker,"Wavelets and their Scientific Applications", Chamman and Hall/CRC, © 1999.

[15] Daubechies, "Ten lectures on wavelets," society for industrial and Applied mathematics, 1992.

[16] Nikhil Rao,"Speech compression using wavelets", ELEC 4801 THESIS PROGEC, School of Information Technology and Electrical Engineering, The university Of Queensland, October 2001.

[17] H.Hermansky, ,"Perceputal Linear Predictive (PLP) analysis of speech", J. Acoust. Soc. Am., 87(4):1738-1752, 1990.

[18] H. Hermansky, ,N. Morgan, "Rasta Processing of Speech", IEEE Trans. on Speech and Audio Proc., Vol.2, No.4, 1994.

[19] M.A.Anusuya and S.K.Katti, " Kannada speech recognition using Discrete Wavelet Transform-PCA", International conference on computer applications-2010, Dec.24-27,Pondicherry,India

[20] M.A.Anusuya and S.K.Katti, " Mel-frequency discrete wavelet coefficients for kannada speech recognition using PCA", International conference on Advances in computer science,Dec.21-22,2010,Trivandrum,Kerala,India.