EURASIP Journal on
Audio, Speech, and Music Processing
a SpringerOpen Journal

RESEARCH   **Open Access**

DWT and LPC based feature extraction methods for isolated word recognition

Navnath S Nehe[1*] and Raghunath S Holambe[2]

Abstract

In this article, new feature extraction methods, which utilize wavelet decomposition and reduced order linear predictive coding (LPC) coefficients, have been proposed for speech recognition. The coefficients have been derived from the speech frames decomposed using discrete wavelet transform. LPC coefficients derived from subband decomposition (abbreviated as WLPC) of speech frame provide better representation than modeling the frame directly. The WLPC coefficients have been further normalized in cepstrum domain to get new set of features denoted as wavelet subband cepstral mean normalized features. The proposed approaches provide effective (better recognition rate), efficient (reduced feature vector dimension), and noise robust features. The performance of these techniques have been evaluated on the TI-46 isolated word database and own created Marathi digits database in a white noise environment using the continuous density hidden Markov model. The experimental results also show the superiority of the proposed techniques over the conventional methods like linear predictive cepstral coefficients, Mel-frequency cepstral coefficients, spectral subtraction, and cepstral mean normalization in presence of additive white Gaussian noise.

Keywords: feature extraction, linear predictive coding, discrete wavelet transform, cepstral mean normalization, hidden Markov model

1. Introduction

A speech recognition system has two major components, namely, feature extraction and classification. Feature extraction method plays a vital role in speech recognition task. There are two dominant approaches of acoustic measurement. First is a temporal domain or parametric approach such as linear prediction [1], which is developed to closely match the resonant structure of human vocal tract that produces the corresponding sound. Linear prediction coefficients (LPC) technique is not suitable for representing speech because it assumes signal stationary within a given frame and hence not analyze the localized events accurately. Also it is not able to capture the unvoiced and nasalized sounds properly [2]. Second approach is nonparametric frequency domain approach based on human auditory perception system and known as Mel-frequency cepstral coefficients (MFCC) [3]. The widespread use of the MFCCs is due

to its low computational complexity and better performance for ASR under clean matched conditions. Performance of MFCC degrades rapidly in presence of noise and degradation is directly proportional to signal-to-noise ratio (SNR). Poor performance of LPC and its different forms like reflection coefficients, linear prediction cepstral coefficients (LPCC) as well as MFCC and its various forms [4] in noisy conditions has led many researchers to investigate alternative robust feature extraction algorithms.

In the literature, various techniques have been proposed to improve the performance of ASR systems in the presence of noise. Speech enhancement techniques such as spectral subtraction (SS) [5] or cepstrums from difference of power spectrum [6] reduce the effect of noise either using statistical information of noise or filtering the noise from noisy speech before feature extraction. Techniques like perceptual linear prediction [7] and relative spectra [8] incorporate some of the features of the human auditory mechanism and give noise robust ASR. Feature enhancement techniques like cepstral mean subtraction [9] and parallel model combination

* Correspondence: nsnehe@yahoo.com
[1]Department of Instrumentation Engineering, Pravara Rural Engineering College, Loni 413736, Maharashtra, India
Full list of author information is available at the end of the article

Springer

[10] improve ASR performance by compensating for mismatch effects in cepstral domain features.

In another approach [11-16] wavelet transform and wavelet packet tree have been used for speech feature extraction in which the energies of wavelet decomposed subbands have been used in place of Mel filtered subband energies. Because of its better energy compaction property [17], wavelet transform-based features give better recognition accuracy than LPC and MFCC. Mel filter-like admissible wavelet packet structure [14] performs better than MFCC in unvoiced phoneme recognition. Wavelet subband features proposed in [15] used normalized subband energies as features which show good performance in presence of additive white noise. However, in these wavelet-based approaches, the time information is lost due to use of wavelet subband energies. We used the actual wavelet coefficients proposed in [18], which preserve the time information, and also these features performed better than LPCC and MFCC due to the combined advantages of LPC and WT. LPC can better distinguish words having distinct vowel sounds [19] and WT can model the details of unvoiced sound portions of speech signal. However, the performance of these features is not well for the noisy speech recognition.

We propose the modification in the features proposed in [18] to derive effective, efficient, and noise robust features from the frequency subbands of the frame. Each frame of speech signal is decomposed (uniformly/dyadic) into different frequency subbands using discrete wavelet transform (DWT) and each subband is further modeled using linear predictive coding (LPC). The WT has a better capability to model the details of unvoiced sound portions. Hence, the subband decomposition has been performed by means of DWT. DWT is more popular in the field of digital signal processing due to its multiresolution capability and it has the property of constant Q, which is one of the demands of many signal processing applications, especially in the processing of the speech signals (as human's hearing system is constant Q perceptional) [20]. Wavelet decomposition results in a logarithmic set of bandwidths, which is very similar to the response of human ear to frequencies (logarithmic fashion). The LPC coefficients derived from the speech subbands obtained after DWT decomposition provide WLPC features [18]. Further these features were normalized in cepstrum domain using well-known cepstrum mean normalization (CMN) technique to get the noise robust features. These new features are denoted as wavelet subband-based cepstral mean normalized features (WSCMN) which perform better in additive white noise environment. The performance of the proposed features is tested on TI-46 and Marathi digits database using continuous density hidden Markov model (CDHMM) as a classifier.

The rest of the article is organized as follows. In Section 2, we describe a brief theory about DWT. The proposed WLPC feature extraction and its normalization are described in Section 3. The various experiments and recognition results are given in Section 4. Section 5 gives the concluding remarks based on the experimentation.

2. Discrete wavelet transform

The speech is a nonstationary signal. The Fourier transform (FT) is not suitable for the analysis of such nonstationary signal because it provides only the frequency information of signal but does not provide the information about at what time which frequency is present. The windowed short-time FT (STFT) provides the temporal information about the frequency content of signal. A drawback of the STFT is its fixed time resolution due to fixed window length. The WT, with its flexible time-frequency window, is an appropriate tool for the analysis of nonstationary signals like speech which have both short high frequency bursts and long quasi-stationary components also.

WT decomposes signals over translated and dilated mother wavelets. Mother wavelet is a time function with finite energy and fast decay. The different versions of the single wavelet are orthogonal to each other. The continuous wavelet transform (CWT) is given by Equation (1) where the function $\psi(t)$, a, and b are called the (mother) wavelet, scaling factor, and translation parameter, respectively.

$$W_x(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-b}{a} \right) dt. \qquad (1)$$

As CWT is a function of two parameters, it contains high redundancy while analyzing the signals. Instead of this, analysis of the signal using small number of scales with varying number of translations at each scale, i.e., discretizing scale and translation parameters as $a = 2^j$ and $b = 2^j k$ gives DWT. DWT theory [20,21] requires two sets of related functions called scaling function and wavelet function given by

$$\phi(t) = \sum_{n=0}^{N-1} h[n]\sqrt{2}\phi(2t - n) \qquad (2)$$

and

$$\psi(t) = \sum_{n=0}^{N-1} g[n]\sqrt{2}\phi(2t - n), \qquad (3)$$

where function $\varphi(t)$ is called scaling function, $h[n]$ is an impulse response of a low-pass filter, and $g[n]$ is an impulse response of a high-pass filter. The scaling and wavelet functions can be implemented effectively using a pair of filters, i.e., $h[n]$ and $g[n]$. These filters are called a quadrature mirror filters that satisfy the property $g[n] = (-1)^{1-n} h[1-n]$ [17]. The input signal is low-pass filtered to give the approximate components and high-pass filtered to give the detail components of the input speech signal. The approximate signal at each stage is further decomposed using same low-pass and high-pass filters to get the approximate and detail components for the next stage. This type of decomposition is called dyadic decomposition, whereas decomposition of detail signal along with the approximate signal at each stage is called uniform decomposition. Dyadic decomposition divides the input signal bandwidth into the logarithmic set of bandwidths, whereas the uniform decomposition divides it into the uniform set of bandwidths.

In speech signal, high frequencies are present very briefly at the onset of a sound while lower frequencies are present latter for long period [21]. DWT resolves all these frequencies well. The DWT parameters contain the information of different frequency scales. This helps in getting the speech information of corresponding frequency band. In order to parameterize the speech signal, the signal is decomposed into four frequency bands uniformly or in dyadic fashion.

3. Proposed WLPC feature extraction

Among the speech recognition approaches, the family based on LPC coefficient and their cepstrum (LPCC) is well known for its performance and relative simplicity. LPC are the coefficients of an auto-regressive model [2] of a speech frame. The all-pole representation of the vocal tract transfer function is as given below

$$H(z) = \frac{G}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{4}$$

where a_p are the prediction coefficients and G is the gain. These LPC can be derived by minimizing the mean square error between the actual samples of speech frame and the estimated samples by autocorrelation method. LPCC were obtained directly using Equation (5) [2].

$$LPCC_i = a_i + \sum_{k=1}^{i-1} \left(\frac{k-i}{i} \right) LPCC_{i-k} a_k \tag{5}$$

where $i = 1,2,...,p$. The obtained LPC and LPCC features cannot capture the high frequency peaks present

in the speech signal and also cannot analyze the localized events accurately which wavelet transform can analyze. However, LPC can better distinguish between the words that have distinct vowel sounds than those share common vowel sounds [19]. WT is able to model the details of unvoiced sound portion of speech than LPC [19]. Also subband signals (wavelet coefficients) obtained from the wavelet decomposition can preserve the time information [12] and LPC can be estimated from such time domain signals easily. So, we can apply LPC technique on each subband signal after the wavelet decomposition which gives the combined benefits of LPC and WT. Hence, the combination of LPC with WT has been proposed in this article.

The LPCC features have been estimated from the subband signals obtained from the DWT in the proposed feature extraction technique. Figure 1 shows the block diagrams of proposed feature extraction systems. Three levels DWT decomposition of preprocessed and windowed speech frames has been done using Daubechies's wavelet filters. Actual wavelet coefficients retain the time information; hence, LPC features have been estimated from the DWT coefficients in time domain. LPC features of pth order have been extracted from each subband of wavelet decomposed speech signal. The schematic of this technique is shown in Figure 1 a. The LPC coefficients obtained from each subband are concatenated to form a final feature vector denoted as Dyadic wavelet decomposed LPC (DWLPC). Thus, the feature vector f_i derived from frame i can be expressed as
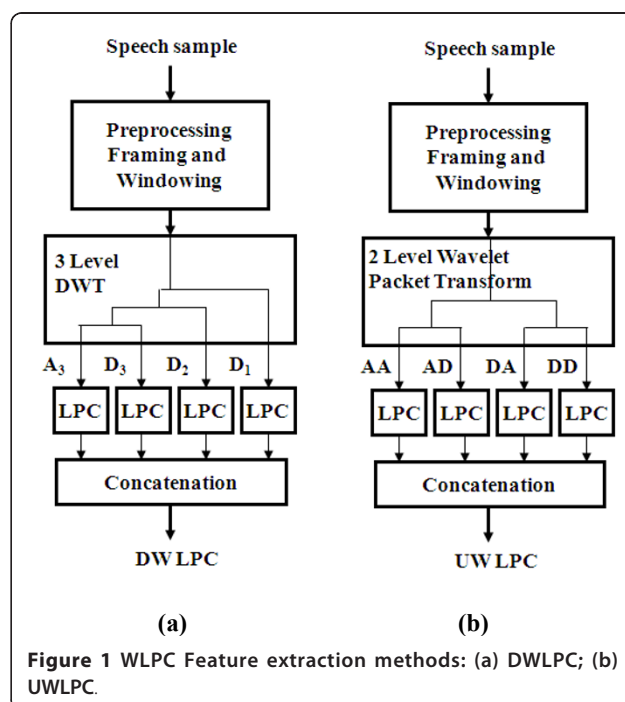

Figure 1 WLPC Feature extraction methods: (a) DWLPC; (b) UWLPC.

$$f_i = [\, a_{A_3} \; a_{D_3} \; a_{D_2} \; a_{D_1} \,]^T , \tag{6}$$

where, a_{A_3} is a row vector formed using prediction coefficients obtained from the approximate components A_3 at third level and a_{D_j} is row vector formed using prediction coefficients obtained from the detail components D_j ($j = 1,2,3$) at jth level. T indicates a vector transpose.

Figure 1b shows the schematic of uniform wavelet decomposed LPC (UWLPC) feature extraction from subbands of uniform bandwidth. The subbands are obtained by two-level wavelet packet decomposition [21]. Then, the UWLPC feature vector is formed similar to DWLPC by concatenation of LPC coefficients estimated from the uniformly decomposed subband signals.

3.1. WSCMN features

CMN [9] is the simplest feature normalization technique to implement. It provides many of the benefits available in the more-advanced normalization algorithms. The LPCC cepstrums were derived using Equation (5) from the WLPC features estimated from the subband signals of each frame. Thus, a sequence of cepstral vectors {x_1, x_2,...,x_T} is obtained from a speech sample. Further these cepstral vectors were normalized using CMN. In its basic form, CMN consists of subtracting the mean feature vector μ_x from each vector x_t and normalizing by variance σ_x to obtain the normalized vector $\hat{x}_t$.

$$\hat{x}_t = \frac{x_t - \mu_x}{\sigma_x} \tag{7}$$

where

$$\mu_x = \frac{1}{T} \sum_t x_t \quad \text{and} \quad \sigma_x^2 = \frac{1}{T} \sum_{t=1}^{T} \left(x_t^2 - \mu_x^2 \right) \tag{8}$$

This gives the proposed WSCMN feature vectors. Figure 2 shows the WSCMN feature extraction steps where U-WSCMN are the uniform decomposed WSCMN feature vectors and D-WSCMN are the dyadic decomposed WSCMN feature vectors.

After normalization, the mean of the cepstral sequence is zero, and it has a variance of one. This normalization is also called as cepstral mean and variance normalization. The CMN makes the features robust to some linear filtering of the acoustic signal, which might be caused by microphones with different transfer functions, varying distance from user to microphone, the room acoustics, or transmission channels [9].

4. Experimental results

This section evaluates the performance of the proposed techniques on isolated words in presence of stationary white noise using TI-46 and own created Marathi databases.

4.1 Databases

The speech recognition experiments were conducted under clean and noisy conditions using the TI-46 and own created Marathi digit database. The TI-46 Speaker Dependent Isolated Word Corpus [22] has two datasets, namely, TI-20 and TI-ALPHA. The TI-20 vocabulary consists of ten English digits "zero" through "nine" and ten control words "yes, no, erase, rubout, repeat, go, enter, help, stop, and start". The TI-ALPHA subset consists of "a" through "z" English alphabets. In both the subsets, data are collected from eight male and eight female speakers. There are 26 utterances of each word from each speaker out of which 10 were used as training tokens and remaining 16 were used as testing tokens. So, TI-20 subset has total 3200 training samples and 5120 test samples, whereas TI-ALPHA has 4160 training samples and 6656 test samples. All the data samples were digitized with sampling frequency 12.5 kHz.

For Marathi database, data were collected from 56 male and 44 female speakers in a quiet room and discretized with sampling frequency 10 kHz. There are 20 utterances of each word from each speaker recorded in 2 different sessions at an interval of 1 week. In each session, ten utterances of each word from each speaker were recorded. For experiments, the samples recorded in first session were used for training and the samples recorded in second session were used for testing. Thus, this database has total 10,000 training samples and 10,000 test samples. Table 1 shows the English digits and their equivalent Marathi digit pronunciation.

4.2 Experimental setup

The input speech samples are pre-emphasized by a first-order filter with transfer function $H(z) = 1\text{-}0.97z^{-1}$. The pre-emphasized speech data are divided into blocks of 25.6 ms duration with 50% overlap between every adjacent frame. The smooth frequency transitions are ensured using a Hamming window to each frame.
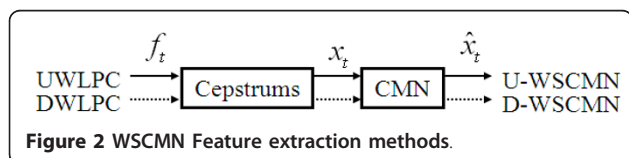

Figure 2 WSCMN Feature extraction methods.

Table 1 English and equivalent Marathi digit pronunciation

Zero	One	Two	Three	Four	Five	Six	Seven	Eight	Nine
Shunya	Ek	Don	Teen	Char	Paach	Saha	Sat	Aath	Nau

Noisy test samples of each dataset (TI-20, TI-ALPHA, and Marathi Digits) were obtained by artificially adding stationary white Gaussian noise under a wide range of SNRs (0, 5, 10, 15, 20, and 30 dB) into the test samples of each dataset. Tests were carried out on clean as well as noisy test samples. For training and testing, diagonal covariance left-right CDHMM [2] with 4-mixtures and 5-states (as this combination yields best performance) was used as a classifier.

4.3 Baseline experiment

The baseline experiments were performed using LPCC and MFCC features on each database. First in the LPCC feature extraction, the prediction coefficients were extracted from each speech frame using 13th order LPC. From the obtained prediction coefficients, cepstral coefficients and its temporal derivatives (first and second derivatives) were extracted and concatenated to form a final LPCC feature vector (this gives feature dimension 39).

In MFCC feature extraction process, the magnitude spectrum of windowed speech frame was filtered using a triangular Mel filter bank consisting of 20 Mel filters. From a set of 20 Mel-scaled log filter bank outputs, MFCC feature vector that consists of 13 MFCC and the corresponding delta and acceleration coefficients (total 39 coefficients) is extracted from each frame. The performance of LPCC and MFCC features was tested on each dataset under clean test condition and presented in Table 2. The recognition results obtained using MFCC features (under clean test condition) are comparable to the state-of-the-art recognition results presented in [23]. These results are used as a baseline for the comparison.

We tested the performance of LPCC and MFCC features for different LPC orders and different number of Mel-filters in the triangular filter bank, respectively. It was observed that 13th-order LPC (p = 13), 20 Mel-filters in filter bank, and feature vector of length 39 (13 LPC/MFCC coefficients and their first and second derivatives) yield best performance on the databases. Hence, the results were obtained for these values of parameters.

4.4 WLPC features

In this section, features were extracted using proposed techniques. In the first type, each speech frame was decomposed into subbands of logarithmic bandwidth by three level DWT and 32nd-order Daubechies's wavelet (the algorithms were tested for various orders and it is observed that 32nd order gives the best performance). Prediction coefficients with different LPC orders (varying from 3 to 7) were derived from the subbands. These prediction coefficients were then concatenated to form DWLPC feature vector. In the second type, each speech frame was decomposed into subbands of uniform bandwidth by two level wavelet packet transform. Then, the prediction coefficients were estimated from the subbands of uniform decomposition similar to first type and were concatenated to form UWLPC feature vector. In both the feature extraction types, we select LPC of order 5 (as it gives the best performance). Five prediction coefficients from each subband give feature vector of dimension 20. Performances of these features were tested using CDHMM with 4-mixtures and 5-states. For the comparison of performance based on the feature dimension, we also considered the 21 coefficients in LPCC and MFCC feature vectors (7 LPC/MFCC coefficients and their first and second derivatives). The performances of LPCC, MFCC, and WLPC (UWLPC/DWLPC) features have been tested on TI-20 database and presented in Table 3.

Percentage recognition rate using LPCC and WLPC (UWLPC/DWLPC) features for different LPC order were also estimated and presented in Figure 3. These results prove that the performance of WLPC (UWLPC/DWLPC) is better than LPCC and MFCC features with half the feature vector length than LPCC and MFCC because the proposed features combine the advantage of identification capability of LPC for vowel and the wavelet's better modeling capability of unvoiced sound portions and high frequency picks of speech sound. Among the WLPC features, DWLPC is superior to UWLPC because the dyadic decomposition in DWLPC mimics the human auditory perception system better.

The performance of MFCC and WLPC (UWLPC and DWLPC) features on TI-Alpha database has been presented in Table 4.
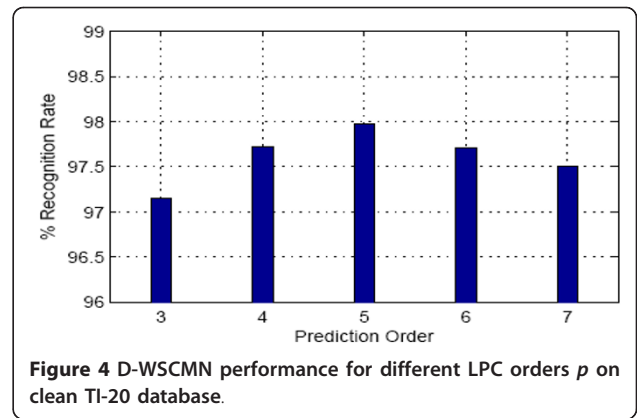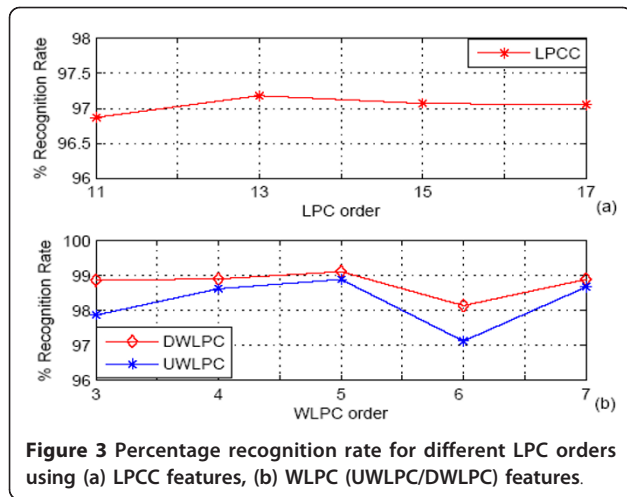
Further, the robustness of the proposed features has been tested by normalizing the features using CMN.

Table 2 Percentage recognition rate of LPCC and MFCC features on various datasets

Dataset	% Recognition rate	
	LPCC	MFCC
TI-20	97.2	98.2
TI-ALPHA	80.6	85.8
Marathi Digits	78.9	84.5

Table 3 Percentage recognition rates of different features on TI-20 database.

Features	Feature vector length	% Recognition rate
LPCC	39	97.2
	21	92.9
MFCC	39	98.2
	21	96.2
UWLPC	20	98.9
DWLPC	20	99.1

Figure 3 Percentage recognition rate for different LPC orders using (a) LPCC features, (b) WLPC (UWLPC/DWLPC) features.

Figure 4 D-WSCMN performance for different LPC orders *p* on clean TI-20 database.

The CMN is applied on the WLPC to get the noise robust WSCMN (D-WSCMN and U-WSCMN) features for the isolated word recognition. The performance of the D-WSCMN for different prediction orders (p) was tested on clean TI-20 database and is presented in Figure 4. From these results it is clear that the D-WSCMN yield better results for $p = 5$. The robustness of WSCMN features was tested on noisy samples generated by adding white Gaussian noise (of SNR 0, 5, 10, and 20 dB) to the test samples of TI-20 dataset. The results of WSCMN features were compared with LPCC, MFCC, SS method [5], and CMN [9] features in Figure 5.

WSCMN feature performance was also tested on clean as well as noisy Marathi digits database. The recognition performance of WSCMN using uniform and dyadic decomposition on this database is shown in Figure 6. It is observed that as compared to MFCC performance on clean data (84.50%), the performance of WSCMN features is significantly increased (100%) on this database. This is because the WSCMN technique is able to capture the difference between the Marathi phonemes more clearly than the MFCC and CMN. Also it gives better performance at various noise levels because of the cepstrum normalization.

5. Conclusions
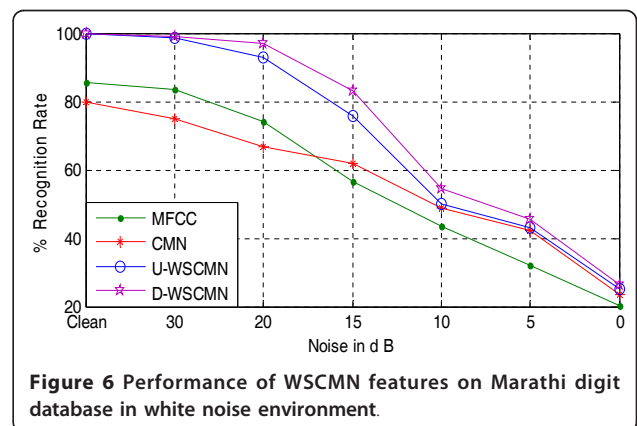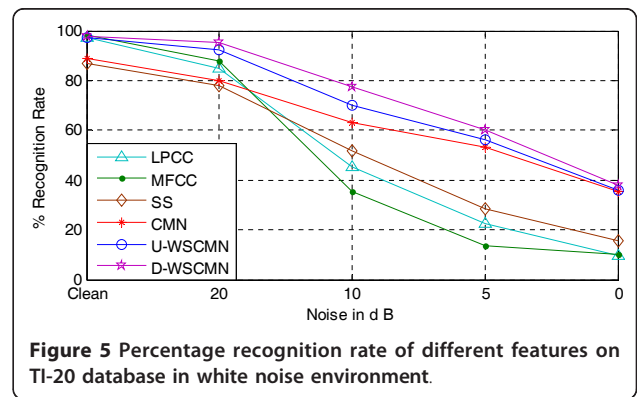In this article, DWT and LPC-based techniques (UWLPC and DWLPC) for isolated word recognition

have been presented. Experimental results show that the proposed WLPC (UWLPC and DWLPC) features are effective and efficient as compared to LPCC and MFCC because it takes the combined advantages of LPC and DWT while estimating the features. Feature vector dimension for WLPC is almost half of the LPCC and MFCC. This reduces the memory requirement and the computational time. It is also observed that the performance of DWLPC is better than UWLPC. This is because the dyadic (logarithmic) frequency

Figure 5 Percentage recognition rate of different features on TI-20 database in white noise environment.

Figure 6 Performance of WSCMN features on Marathi digit database in white noise environment.

Table 4 Performance of WLPC features on TI-Alpha database

Features	Feature vector length	% Recognition rate
MFCC	39	84.1
UWLPC	20	85.2
DWLPC	20	87.0

decomposition mimics the human auditory perception system better than uniform frequency decomposition.

WSCMN features are noise robust features because of normalization in cepstrum domain. It is observed that the proposed WSCMN features yield better performance as compared to the popular existing methods in presence of white noise because this technique is able to capture the difference between the phonemes (especially in Marathi database) more clearly than the MFCC and CMN. It has also been proved experimentally that the proposed approaches provide effective (better recognition rate), efficient (reduced feature vector dimension), and robust features.

Author details
[1]Department of Instrumentation Engineering, Pravara Rural Engineering College, Loni 413736, Maharashtra, India [2]S.G.G.S. Institute Engineering & Technology, Vishnupuri, Nanded, Maharashtra, India

Competing interests
The authors declare that they have no competing interests.

References
1. F Itakura, Minimum prediction residual principle applied to speech recognition. IEEE Trans Acoust Speech Signal Proces. **ASSP-23**, 67–72 (1975)
2. L Rabiner, BH Juang, *Fundamentals of Speech Recognition* (Prentice-Hall Inc., Englewood Cliffs, NJ, 1993)
3. SB Davis, P Mermelstein, Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans Acoust Speech Signal Process. **ASSP-28**(4), 357–366 (1980)
4. K Wang, CH Lee, BH Juang, Selective feature extraction via signal decomposition. IEEE Signal Process Lett. **4**, 8–11 (1997). doi:10.1109/97.551687
5. SF Boll, Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans Acoust Speech Signal Process. **27**, 113–120 (1979). doi:10.1109/TASSP.1979.1163209
6. J Xu, G Wei, Noise-robust speech recognition based on difference of power spectrum. Electron Lett. **36**(14), 1247–1248 (2000). doi:10.1049/el:20000848
7. H Hermansky, Perceptual linear predictive (PLP) analysis of speech. J Acoust Soc Am. **87**(4), 1738–1752 (1990). doi:10.1121/1.399423
8. H Hermansky, N Morgan, RASTA processing of speech. IEEE Trans Speech Audio Process. **2**, 578–589 (1994). doi:10.1109/89.326616
9. AE Rosenberg, CH Lee, FK Soong, Cepstral channel normalization techniques for hmm-based speaker verification, in *Proc ICSLP*, Yokohama, Japan, 1835–1838 (1994)
10. MJF Gales, SJ Young, Robust speech recognition using parallel model combination. IEEE Trans Speech Audio Process. **4**, 352–359 (1996). doi:10.1109/89.536929
11. Z Tufekci, JN Gowdy, Feature extraction using discrete wavelet transform for speech recognition, in *IEEE International Conference Southeastcon 2000*, Nashville, TN, USA, 116–123 (April 2000)
12. M Gupta, A Gilbert, Robust speech recognition using wavelet coefficient features, in *Proc IEEE workshop on Automatic Speech Recognition and Understanding (ASRU'01)*, Madonna di Campiglio, Trento, Italy, 445–448 (December 2001)
13. JN Gowdy, Z Tufekci, Mel-scaled discrete wavelet coefficients for speech recognition, in *Proc IEEE Inter Conf Acoustics, speech, and Signal Processing (ICASSP'00)*, vol. 3. Istanbul, Turkey, 1351–1354 (June 2000)
14. O Farooq, S Datta, Mel filter-like admissible wavelet packet structure for speech recognition. IEEE Signal Process Lett. **8**(7), 196–198 (2001). doi:10.1109/97.928676
15. O Farooq, S Datta, Wavelet based robust sub-band features for phoneme recognition. IEE Vis Image Signal Process. **151**(4), 187–193 (2004)
16. B Kotnik, Z Kačič, A comprehensive noise robust speech parameterization algorithm using wavelet packet decomposition-based denoising and speech feature representation techniques. EURASIP J Adv Signal Process. **1**, 1–20 (2007)
17. S Mallat, *A Wavelet Tour of Signal Processing* (Academic, New York, 1998)
18. NS Nehe, RS Holambe, New feature extraction methods using DWT and LPC for isolated word recognition, in *Proc of IEEE TENCON 2008*, Hyderabad, India, 1–6 (2008)
19. M Krishnan, CP Neophytou, G Prescott, Wavelet transform speech recognition using vector quantization, dynamic time warping and artificial neural networks, in *International Conference On Spoken Language Processing*, Yokohama, Japan, 1191–1193 (1994)
20. Y Hao, X Zhu, A new feature in speech recognition based on wavelet transform, in *Proc IEEE 5th Inter Conf on Signal Processing (WCCC-ICSP 2000)*, vol. 3. Beijing, China, 1526–1529 (21–25 August 2000)
21. KP Soman, KI Ramchandran, *Insight into Wavelets from Theory to Practice*, 2nd edn. (Prentice-Hall of India, New Delhi, 2005)
22. TI 46-Word Speaker-Dependent Isolated Word Corpus, NIST Speech Disc 7-1.1 (1991)
23. DS Pallett, A benchmark for speaker-dependent recognition using the Texas Instruments 20 Word and Alpha-set speech database, in *Proc of Speech Recognition Workshop*, Bristol, UK, 67–72 (1986)