# Chapter 3
# Evaluation of Wavelet-Based Speech Codecs for VoIP Applications

**Manas Ray and Mahesh Chandra**

## 3.1 Introduction

Voice over Internet Protocol (VoIP) is a pioneer technology in the modern communication world that allows delivery of voice calls over packet-switched network like broadband Internet connectivity. This technology utilizes the existing data communication infrastructure to deliver the voice packets. The current challenges faced by VoIP includes relatively high bandwidth requirement, traffic congestion leading to propagation delay, network delay variations, and excessive delay [1]. To minimize these limitations of VoIP, an efficient speech compression technique is desired. The speech signal is required to be compressed so as to conserve the precious resource of bandwidth.

In the recent years, wavelet transforms and its applications are being extensively studied in the field of signal processing. Wavelet transform provides excellent resolution in frequency as well as in time domain [2]. The wavelet transform represents the signal with very high precision and limited storage requirements [3]. The wavelet is defined as limited waveform having zero average value. It is finite in nature. The multi-resolution capability of the wavelet provides us with dilate and translate versions of the wavelet [4]. The resolution of the analysis is determined by the scaling function, and the analysis is performed by the mother wavelet function $\Psi(k)$ [3]. Wavelet transform is calculated by the convolution of original signal $s(k)$ and the mother wavelet function $\Psi(k)$ as defined as follows [5]:

M. Ray (✉) · M. Chandra
Department of Electronics and Communication Engineering, Birla Institute
of Technology Mesra, Ranchi 835215, India
e-mail: manas_ray.me@rediffmail.com

$$W_{\Psi}(m, n) = \int_{-\infty}^{\infty} s(k)\Psi'_{mn}(k)\mathrm{d}k$$

$$= \frac{1}{\sqrt{m}} \int_{-\infty}^{\infty} s(k)\Psi\left(\frac{k-n}{m}\right)\mathrm{d}k \qquad (3.1)$$

where $s(k)$ is the original signal, '$m$' is the scaling factor, '$n$' is the translation parameter, and $\Psi(k)$ is the mother wavelet. The wavelet function is given by

$$\Psi_{m,n} = \frac{1}{\sqrt{m}}\Psi\left(\frac{k-n}{m}\right) \qquad (3.2)$$

The discrete version of the continuous wavelet transform (CWT) with dyadic grid parameters of translation $n = p$ and the scale $m = 2^j$ and the mother wavelet is defined by

$$\Psi(x) = 2^{j/2}\Psi(2^j x - p) \qquad (3.3)$$

Similarly, the scaling function is defined as follows:

$$\phi(x) = 2^{j/2}\phi(2^j x - p) \qquad (3.4)$$

The original function $f(x)$ can be obtained from the scaling and the wavelet functions from the [6]:

$$f(x) = \sum_{p=-\infty}^{\infty} c_p\phi_p(x) + \sum_{p=-\infty}^{\infty} d_{j,p}\Psi_{j,p}(x) \qquad (3.5)$$

where $C_p$ are the average coefficients, and $d_{j,p}$ are detail coefficients.

## 3.2 Various Wavelet Families

In this paper, we evaluate the following wavelet families Haar, Daubechies, Discrete approximation of Meyer wavelet (dmey) and Coiflets. Each of these wavelet families is defined as follows:

(a) Haar Wavelet (Haar)

Haar wavelet is the simplest possible wavelet [7]. For a signal represented by $2^t$ values, the wavelet transform recursively provides the difference and forwards the sum to the next level, resulting in $2^t - 1$ differences and one total summation. It is not continuous. The wavelet function $\Psi(x)$ is defined as follows

$$\Psi(x) = \begin{cases} 1 & 0 \le t \le \frac{1}{2} \\ -1 & \frac{1}{2} \le t \le 1 \\ 0 & \text{Otherwise} \end{cases} \tag{3.6}$$

The scaling function is defined as follows

$$\Phi(x) = \begin{cases} 1, & 0 \le t < 1 \\ 0, & \text{Otherwise} \end{cases} \tag{3.7}$$

(b) Daubechies Wavelet

Daubechies wavelets are orthogonal wavelets, having largest number of vanishing moments for some support and are commonly used for the analysis of a signal. Here, the scaling and the wavelet functions are not defined [8]. The number of coefficients generated is defined by the index $N$ of the coefficients, and the number of vanishing moments is $N/2$ [3].

(c) Discrete Approximation of Meyer Wavelet (dmey)

The discrete format of the Meyer wavelet function is defined as follows

$$G_o\left(e^{j\omega}\right)\sqrt{2}\sum_K \Phi(2\omega + 4k\Pi) \tag{3.8}$$

Given the basis function '$\Phi$', DTFT techniques are employed to obtain the scale coefficients [9].

(d) Coiflet Wavelet

Coiflets are wavelets having scaling functions with vanishing moments. The wavelet is near symmetric and has $N/3$ vanishing moments, and scaling function has $N/3 - 1$ vanishing moments [3]. If the taps $N = 6p$, then $2p$ number of vanishing moment conditions are imposed on wavelet function and $2p - 1$ on scaling function and the remaining on normality and orthogonality conditions.

Thus, the conditions imposed are as follows [10]:

$$\int \phi(k)\mathrm{d}k = 1 \tag{3.9}$$

$$\int \phi(k)\phi(k - l)\mathrm{d}k = \delta_{0,l} \tag{3.10}$$

$$\int k^n \Psi(k)\mathrm{d}k = 0 \quad \text{for } n = 0, 1, 2, \ldots 2p - 1 \tag{3.11}$$

$$\int k^n \phi(k)\mathrm{d}k = 0 \quad \text{for } n = 0, 1, 2, \ldots 2p - 1 \tag{3.12}$$

## 3.3  Speech Signal Processing Using Wavelet Transform

The speech signal processing or compression by wavelet transform is performed by choosing a particular wavelet function. The speech quality requirements of the codec govern the selection of the wavelet function for the analysis. The objective of the processing is to maximize the signal quality and minimize reconstructed error variance [11]. Wavelets decompose a signal into components of different frequency bands called as resolution. The signal compression is achieved by reconstructing the signal by considering a limited set of approximation coefficients and some detail coefficients. This is done by the process of thresholding, wherein coefficients falling below a threshold value are ignored and made equal to zero [11]. The signal is reconstructed by performing inverse wavelet transform using the coefficient values which are above the threshold values. Generally, 5-level decomposition is adequate for speech signals [12]. Figure 3.1 [13] shows the process of the speech signal processing for the purpose of compression using the wavelet transform technique.

## 3.4  Performance Evaluation Parameters

The speech codecs based on the above-defined families of wavelets are implemented in MATLAB for the simulation purpose. The acceptability of the performance of the wavelet-based speech codec for VoIP application is gauged by the subjective testing of mean opinion score (MOS), wherein the original signal and re-constructed signal are presented to a user, who then provide a performance rating between 1 and 5, where 5 is excellent grade [14]. Further, the performance evaluation of the wavelet-based codec is carried out by objective testing of the speech samples. The tests were carried out by comparing the performance in terms of compression ratio (CR), SNR, NRMSE [12, 13], and retained signal energy (RSE). The expressions of these parameters are given below.

$$CR = \frac{\text{Length of } (o(k))}{\text{Length of } (p(k))} \tag{3.13}$$

where

$o(k)$   is the input signal
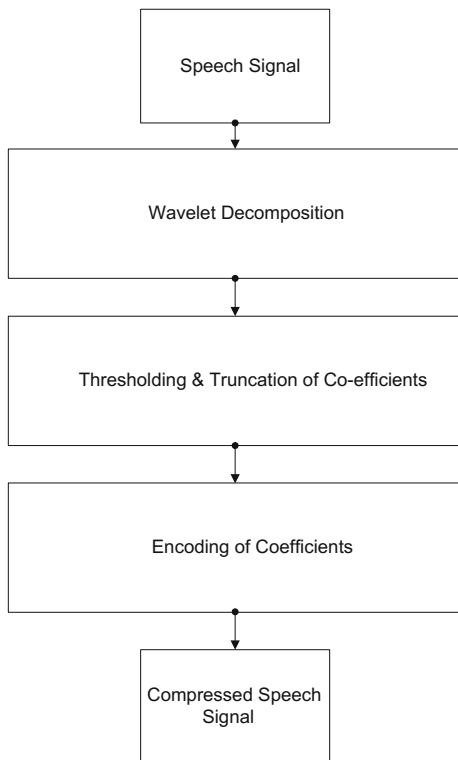$p(k)$   is the re-constructed signals, respectively

$$SNR = 10 \log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) \tag{3.14}$$

where $\sigma_x^2$ and $\sigma_e^2$ are mean square of the input signal and the mean square difference between the input and re-constructed signal, respectively.

Normalized root mean square error (NRMSE) is given by

**Fig. 3.1** Speech compression
using DWT



$$\text{NRMSE} = \sqrt{\frac{(o(n) - p(n))^2}{(o(n) - \mu o(n))^2}} \tag{3.15}$$

where

$o(n)$     is the original input signal,
$p(n)$     is the signal, re-constructed and
$\mu o(n)$   is the mean of the original signal.

Retained signal energy (RSE) [15] is defined as follows

$$\text{RSE}\,(\%) = \frac{\|o(n)\|^2}{\|p(n)\|^2} \times 100 \tag{3.16}$$

where

$\|o(n)\|$   is the original signal norm
$\|p(n)\|$   is the norm of the re-constructed signal.

**Table 3.1** Details of sample sentences used in the experiment

| Sample no. | Language | Speaker |
|---|---|---|
| 1 | Hindi | Male |
| 2 | Hindi | Male |
| 3 | English | Male |
| 4 | English | Female |
| 5 | Hindi | Female |
| 6 | Hindi | Female |

## 3.5 Results

Discrete wavelet transform-based codec is simulated in MATLAB based on the speech compression principle adopted in wavelet transforms. The test sentences as presented in Table 3.1 are iterated against each of the set of 4 different wavelet families, viz. Haar, Daubechies, dmey, and Coiflet wavelets.

The speech signal is decomposed into 5-level approximation and detail coefficients. A global threshold value is used for the decomposition of signal. The quality of the signal was measured based on MOS, SNR, RSE, and compression ratio.

The results are shown in the following figures. Figure 3.2 shows the comparison of the wavelets in terms of the MOS, Fig. 3.3 compares the wavelets in terms of the compression ratio, Fig. 3.4 compares the wavelets in terms of SNR, and Fig. 3.5 shows the comparison in terms of the RSE %.
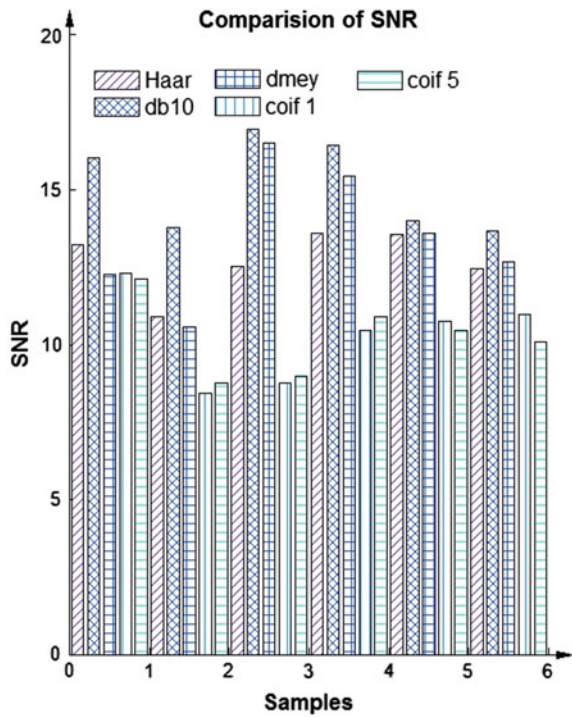
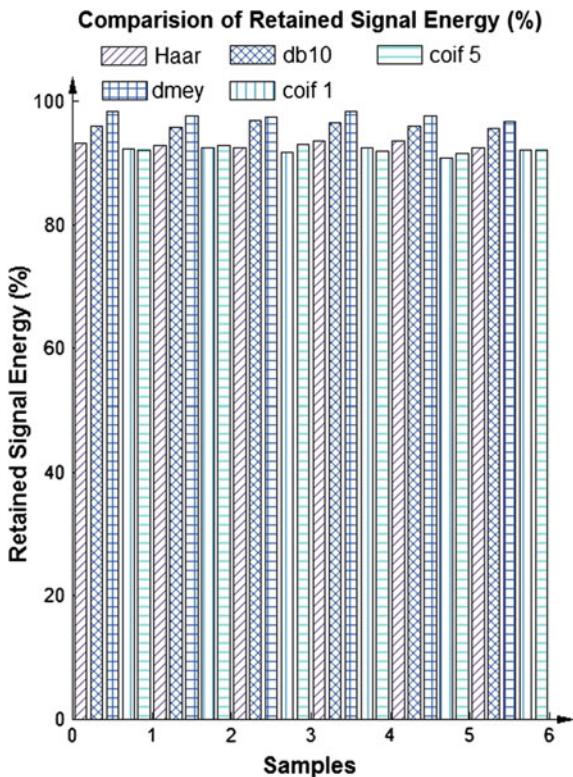**Fig. 3.2** Comparison of MOS of wavelets
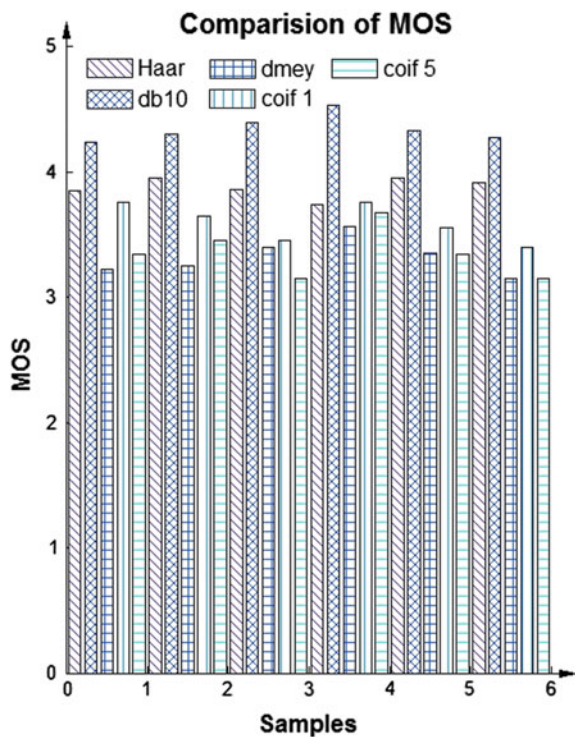
**Fig. 3.3** Comparison of compression ratio of wavelets
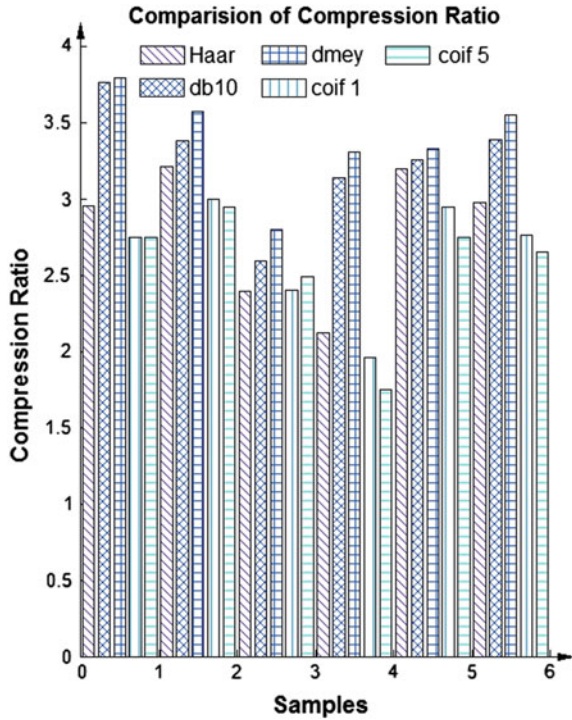


**Fig. 3.4** Comparison of SNR of wavelets

**Fig. 3.5** Comparison of
retained energy



It can be seen from the above that dmey wavelets provide excellent results in terms of % energy retention and compression ratio, followed by Daubechies family. Further wavelets of Daubechies family provide better degree of performance in terms of the MOS and SNR of the signal.

## 3.6 Conclusions

Wavelet-based speech coding, in general, offers a good degree of compression of the speech signal, whose magnitude can be varied easily. The Haar wavelet transform is the straight forward and fastest transform to be used for speech compression. However, due to its discontinuity, it is not advantageous for the simulation of speech signals. Daubechies wavelet has shown its superiority over other families of wavelet for speech compression in terms of all parameters such as % compression and SNR value and hence extensively used in various speech processing applications. Further, it can be inferred from the above that the average MOS of wavelet-based speech codec is in the range of 3.9–4.5, which is near toll

quality; hence, they compare well with the currently deployed speech codec in the VoIP applications. The results further reveal that the performance of codec under study remains unaffected with change in language or speakers.

# References

1. Ray M, Chandra M, Patil BP (2012) Evaluation of CDMA microwave links at different environments for VoIP applications. Int J Adv Res Comput Commun Eng (IJARCCE) 1 (8):508–512
2. Karam J (2007) Various speech processing techniques for speech compression and recognition. In: Proceeding of world academy of science, engineering and technology, vol 26, pp 209–213, ISSN 1307-6884
3. Somani KP, Ramchandran KI, Resmi NG (2009) Insights into wavelets from theory to practice. Prentice Hill International
4. Daubechies I (1998) Ortho-normal bases of compactly supported wavelets. IEEE Commun Pure Appl Math 41:906–996
5. Debdas S, Jagrit V, Chandrakar C, Quereshi MF (2011) Application of wavelet transform for speech processing. Int J Eng Sci Technol (IJEST) 3(8):6666–6670
6. Agbinya JI (1996) Discrete wavelet transform techniques in speech processing. IEEE TENCON-Digital Signal Process Appl 514–519
7. http://en.wikipedia.org/wiki/Haar_wavelet
8. http://en.wikipedia.org/wiki/Daubechies_wavelet
9. Kornsing S, Srinonchat J (2012) Enhancement speech compression technique using modern wavelet transforms. IEEE Int Symp Comput Consum Control 393–396
10. http://en.wikipedia.org/wiki/Coiflet
11. Kinsner W, Langi A (1993) Speech and image signal compression using wavelets. In: IEEE Wescanex Conference Proceedings, pp 368–375
12. Ambika D, Radha V (2012) A comparative study between discrete wavelet transform and linear predictive coding. In: IEEE world congress on information and communication technologies, pp 965–969
13. Joseph SM (2010) Spoken digit compression using wavelet. IEEE international conference on signal and image processing, pp 255–259
14. Ray AK, Acharya T (2004) Information technology: principles and applications. Prentice-Hill of India Private Limited, New Delhi
15. Nan L, China J (2009) A new adaptive threshold algorithm to speech enhancement based on minimum description length criterria. In: IEEE international conference on information engineering and computer science, pp 1–9