International Journal of Engineering Research & Management Technology ISSN: 2348-4039
March-2016 Volume 3, Issue-2
Email: editor@ijermt.org www.ijermt.org

Speech Compression Using Slant Let Transform Based On Different Quantization

Manish Kumar Sahu
Shri Ram Institute of Technology
Jabalpur, Madhya Pradesh

Ravi Mohan*Divyanshu Rao**
Profacer* Asstient Profacer**
Shri Ram Institute of Technology
Jabalpur, Madhya Pradesh

1.1 INTRODUCTION:

Motive of speech signal compression is to reducing the transmission bit rate, storage capacity by removing the redundancies present in speech signal while keeping the hearing quality and lucidity of speech signal preserved. This chapter presents new method for speech signals compression using slantlet transform techniques with scalar and vector quantization methods and huffman encoding. The performance of the implemented method is evaluated based on signal to noise ratio, peak signal to noise ratio, normalized root mean square error and compression ratio and tested on 4 kHz 8-bit speech signals. The simulation results of developed speech compression technique not only give the better compression ratio but also yield good fidelity parameters as compared to other transforms method.

1.2 SLANTLET TRANSFORM:

Slantlet transform (SLT), recently developed multiresolution technique is an orthogonal discrete wavelet transform with two zero moments and with improved time localization. SLT well suited for piecewise linear data. It also preserves the basic characteristics of the usual filter-bank such as octave band characteristic and a scale dilation factor of two [34-37]. SLT is based on the principle of designing different filters for different scales unlike iterated filter bank approaches for the DWT. SLT is basically an orthogonal DWT and the construction of the slantlet is based on a filter bank structure where different filters are used for each scale. Selesnick [34] proposed SLT in 1999, the filterbank defined by Selesnick for slantlet transform works a similar parallel structure like DWT providing exactly a scale dilation factor of two. SLT can also be applied in medical image processing for the classification of magnetic resonance images (MRI). SLT makes use of a special class of bases, which is constructed using Gram-Schmidt orthogonalization procedure [38].
A two-scale iterated DWT filter bank shown in Fig. 1.1 and its equivalent form in Fig. 1.2. SLT filterbank works the structure of the equivalent form shown in Fig. 1.2, but it is occupied by different filters that are not products. With this extra degree of freedom obtained by giving up the product form, filters of shorter length are designed satisfying orthogonality and zero moment conditions [36]. In case of two channel Daubechies filter is the shortest filter which makes the filter bank orthogonal and has K zero moments. When K=2 zero moments the filters of Fig.1.2 are of lengths 10 and 4 but in case of slantlet filter bank with K=2 zero moments the filter has lengths 8 and 4 which is shown in Fig. 1.2. Thus in case of two-scale slantlet filter bank has a filter length which is two samples less than that of a two-scale iterated Daubechies-2 filter bank. This difference will increases with the increased number of stages. Since there is no tree structure for slantlet it can be efficiently implemented like an iterated DWT filter bank. The computational complexities of the SLT are of the same as that of the DWT, but SLT gives better performance in denoising and compression of the signals.

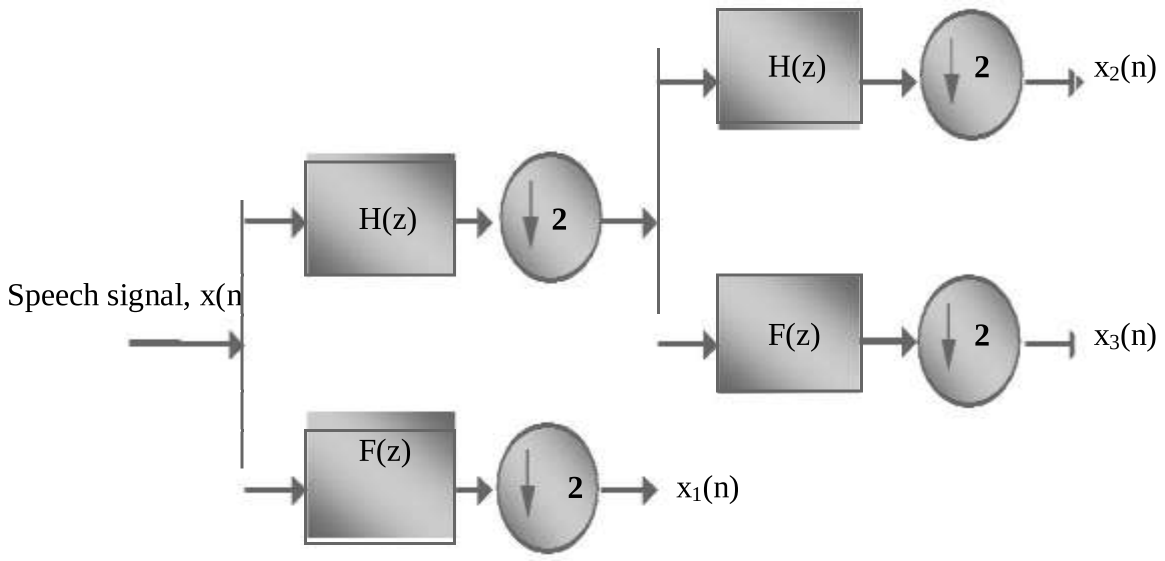

Fig. 1.1: 2-scale slantlet filter bank [34].

The slantlet filter bank consist of $g_i(n)$, and $f_i(n)$ filter, where i denote the scale. The length of the filters for scale i will be proportional to 2^i. That is approximately true for iterated filter banks. However, it is exact for slantlet filterbanks. For L-scale filter bank there are 2L channels. The low-pass filter is to be called $h_L(n)$ and filter adjacent to the low-pass filter is to be called as $f_L(n)$.Both low pass filter $h_L(n)$ and high pass filter $f_L(n)$ are to be followed by down sampling by 2^L. And again the remaining 2L-2 channels are filtered by $g_i(n)$ and its shifted time-reverse for i=1,…, L-1. And this $g_i(n)$ and $h_i(n)$ filter is followed by down sampling by 2^{L-1}.

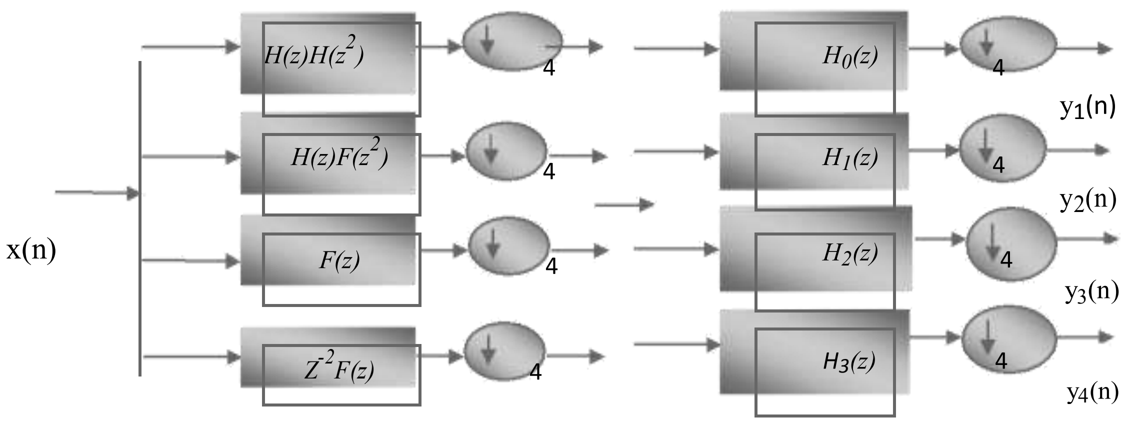

Fig. 1.2: Equivalent structure of slantlet filter bank [34].

Filter $g_i(n)$ can be described by four parameters and can be written as:

$$g_i(n) = \begin{cases} a_0 + a_1 n & \text{for } n = 0, \ldots, 2^i - 1 \\ b_0 + b_1 (n - 2^i) & \text{for } n = 2^i, \ldots, 2^{i+1} - 1 \end{cases} \quad 1.1$$

For calculating $g_i(n)$ calculate a_0 , a_1 , b_0 ,b_1 which is given by the equation

$$m \square 2^i \qquad 1.2$$

$$s_1 \square 6 \sqrt{m / ((m^2 \square 1)(4m^2 \square 1))} \qquad 1.3$$

$$t_1 \square 2 \sqrt{3 / (m .(m^2 \square 1))} \qquad 1.4$$

$$s_0 \square \square \square s_1 .(m \square 1) / 2 \qquad 1.5$$

$$t_0 \square ((m \square 1)s_1 / 3 \square mt_1)(m \square 1) / (2m) \qquad 1.6$$

$$a_0 \square (s_0 \square t_0) / 2 \qquad 1.7$$

$$b_0 \square (s_0 \square t_0) / 2 \qquad 1.8$$

$$a_1 \square (s_1 \square t_1) / 2 \qquad 1.9$$

$$b_1 \square (s_1 \square t_1) / 2$$

$$h_i(n) \square \begin{matrix} a_{01} \square a\, n \\ b_{01} \square b\,(n - 2^i) \end{matrix}$$

$$f_i(n) \square \begin{matrix} c_{01} \square c\, n \\ d_{01} \square d\,(n - 2^i) \end{matrix}$$

Where $m \square 2^i$

$$u \square 1/ \sqrt{m} \qquad 1.13$$

$$v \square \sqrt{(2 m^2 \square 1) / 3} \qquad 1.14$$

$$a_0 \square u.(v \square 1) / (2m) \qquad 1.15$$

$$b_0 \square u .(2 m \square v \square 1) / (2m) \qquad 1.16$$

$$a_1 \square u / m \qquad 1.17$$

$$b_1 \square \square \square a_1 \quad 1.18$$

$$q \square \sqrt{3/(m.(m^2 \square 1))}\ /m \quad 1.19$$

$$c_1 \square \square \square q.(v \square m) \quad 1.20$$

$$d_1 \square \square \square q.(v \square m) \quad 1.21$$

$$d_0 \square d_1.(v \square 1 \square 2m)/2 \quad 1.22$$

$$c_o \square c_1.(v \square 1)/2 \quad 1.23$$

1.3 QUANTIZATION:

1.10

$$\text{for } n \square 0,\ldots 2^{i} \square 1 \quad 1.11$$
$$\text{for } n \square 2^{i},\ldots 2^{i \square 1} \square 1$$

$$\text{for } n \square 0,\ldots\ldots 2^{i} \square 1 \quad 1.12$$
$$\text{for } n \square 2^{i},\ldots 2^{i \square 1} \square 1$$

The transform data can be quantized by two methods like scalar quantization and vector quantization. In scalar quantization process, the transform coefficients are quantized using uniform step size, which depends on three parameters maximum (M_{max}) and minimum (M_{min}) values in the signal matrix, and the number of quantization level (L). Once these parameters are found, and then compute the step size. In this process input signal is divided in to L level and equal interval size ranging from M_{min} to M_{max} and made a quantization table. In a second quantization method i.e. vector quantization quantizing a set of data samples jointly as a vector. Shannon's rate distortion theory says that better results are always obtained when vectors instead of scalars are encoded. Vector quantizes that map vectors in a multidimensional space into a finite set of vector called reproduction vectors and this reproduction vector is called as codebook. This codebook is then further used to encode and decode the data set. Fig 1.3 shows a basic block diagram representation of vector quantizer. In this process codebook is generated by generalized Lloyd

(GL) algorithm, also known as the LBG (Linde, Buzo, Gray) algorithm [26

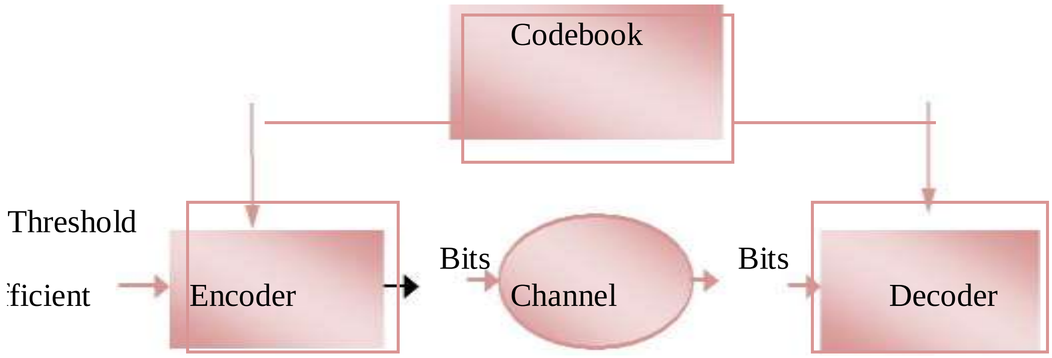

Fig. 1.3: Structure of vector quantizer [26].

1.4 PROPOSED COMPRESSION METHODOLOGY FOR SPEECH:

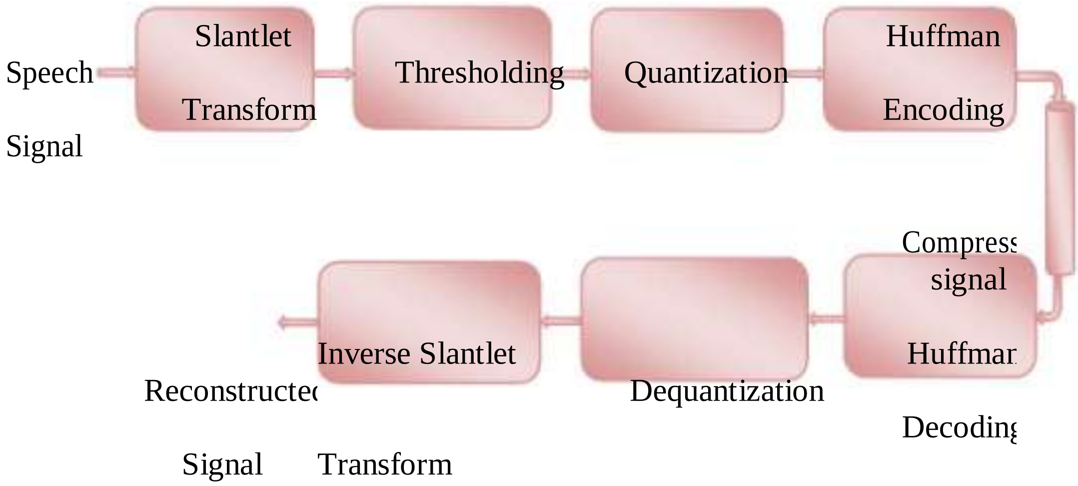

Fig. 1.4: Flowchart for slantlet transform based speech compression method.

Step 1: A two-scale slant let transform is applied to the given speech signal. The speech signal is fed to the two-scale SLT filter bank, where the outputs are down sampled by a factor four. The 54 slant let filter coefficients used in this experiment is obtained from [28]. The filter coefficients of the slantlet transform are $H_0(z)$, $H_1(z)$, $H_2(z)$ and $H_3(z)$. SLT contains most of its input signal energy in some few coefficients and magnitudes of the remaining coefficients are insignificant.
Step 2: After performing SLT on the signal, thresholding is applied on transform coefficients which made a fixed percentage of coefficients equal to zero. There are two types of thresholding. The first type is called as global thresholding; in this method the threshold value is select manually. This value is chosen from transform coefficients (0 ...xjmax), where $x^j_{max}$is the maximum value of coefficient. The second method is known as level thresholding in which, the thresholdvalue is calculated using Birge-Massart strategy [16]. But in this work, we applied global thresholding. Thereon, quantization is performed on the truncated coefficients.

Step 3: In quantization process, threshold coefficients can quantize either using scalar quantization or vector quantization. In scalar quantization process, threshold coefficients are quantized with uniform step size, which depends on three parameters: maximum (M_{max}) and minimum (M_{min}) values in the signal matrix, and the number of quantization level (L).after getting these parameter we calculate the step size by (1.24).

$$\Delta = (M_{max} - M_{min}) / L \qquad 1.24$$

Then, divide the input signal into L+1 level with equal interval size ranging from M_{min} to M_{max} to made the quantization table. When quantization has completed then quantized values are fed to the next stage of compression and apply the huffman encoding on quantized value here we get the compressed signal and transmit the signal through the channel. In a second quantization method i.e. vector quantization, quantizing a set of data samples jointly as a vector. In this process quantize the signal by using codebook which is generated by using generalized Lloyd (GL) algorithm and calculate the indices and then apply huffman encoding. Here we get the compress signal and transmit the signal through the channel.

1.5 RESULTS AND DISCUSSION:

In previous section a new methodology for speech signal compression using slantlet transform and different quantization using huffman encoding has developed. The performance of the developed method is measured by considering fidelity of the reconstructed signal to original signal. For this, following fidelity assessment parameters CR, SNR, PSNR, and NRMSE are considered.

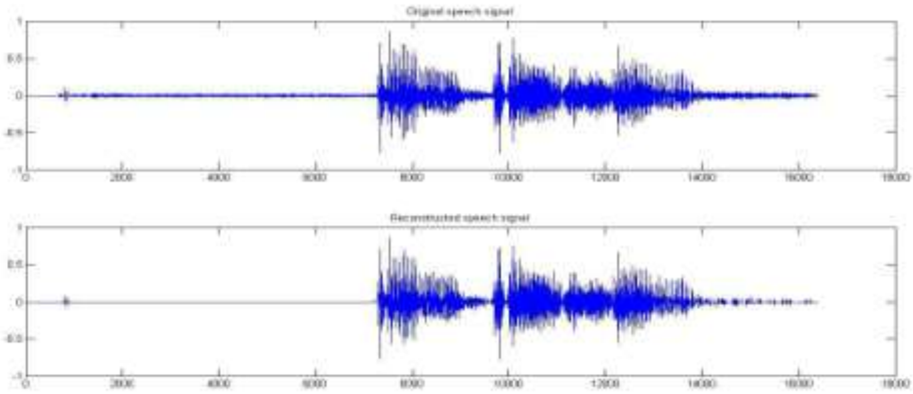

Fig. 1.6: Original speech signal and reconstructed speech signal.

Table 1.1: Fidelity parameters of the proposed speech compression method based on SLT with scalar quantization

Signal	CR	SNR	PSNR	NRMSE
Compression	0.4976	35.8974	70.7377	0.0164
Apple	0.4832	33.5323	72.2907	0.0152
Name	0.4895	37.8174	67.5357	0.0205
Mango	0.4764	36.4615	74.1162	0.0139
Hello	0.4897	34.0294	71.1922	0.0091

Peacock	0.4777	38.3022	77.1835	0.0118
Frequency	0.4828	29.3212	74.5139	0.0125
Country	0.4909	30.1338	75.8538	0.0112

Table 1.2: Fidelity parameters of the proposed speech compression method based on SLT with vector quantization

Signal	CR	SNR	PSNR	NRMSE
Compression	0.3957	40.6579	80.2561	0.0095
Apple	0.3717	37.7944	80.8152	0.0093
Name	0.3934	40.6384	73.1778	0.0088
Mango	0.3844	40.1581	81.4669	0.0091
Hello	0.3827	39.7646	82.6627	0.0047
Peacock	0.3511	42.3451	85.2235	0.0074
Frequency	0.3794	35.1399	86.1513	0.0064
Country	0.3946	35.2283	85.9351	0.0061

Table 1.3: Fidelity parameters of the proposed speech compression method based on different transform with scalar and vector quantization.

Signal	Transform	CR	SNR	PSNR	NRMSE
Compression	DWT+SQ	0.6403	12.4734	30.1954	0.0234
	DWT+VQ	0.4524	23.0932	42.726	0.0114
	SLT+SQ	0.4976	35.8974	70.7377	0.0164
	SLT+VQ	0.3957	40.6579	80.2561	0.0095
Apple	DWT+SQ	0.6311	14.2367	29.8428	0.0319
	DWT+VQ	0.4624	20.0932	38.3994	0.0124
	SLT+SQ	0.4832	33.5323	72.2907	0.0152
	SLT+VQ	0.3717	37.7944	80.815	0.0093
Name	DWT+SQ	0.6311	14.2367	29.8428	0.0329
	DWT+VQ	0.4624	20.0932	38.3994	0.0148
	SLT+SQ	0.4895	37.8174	67.5357	0.0205
	SLT+VQ	0.3934	40.6384	73.1778	0.0124

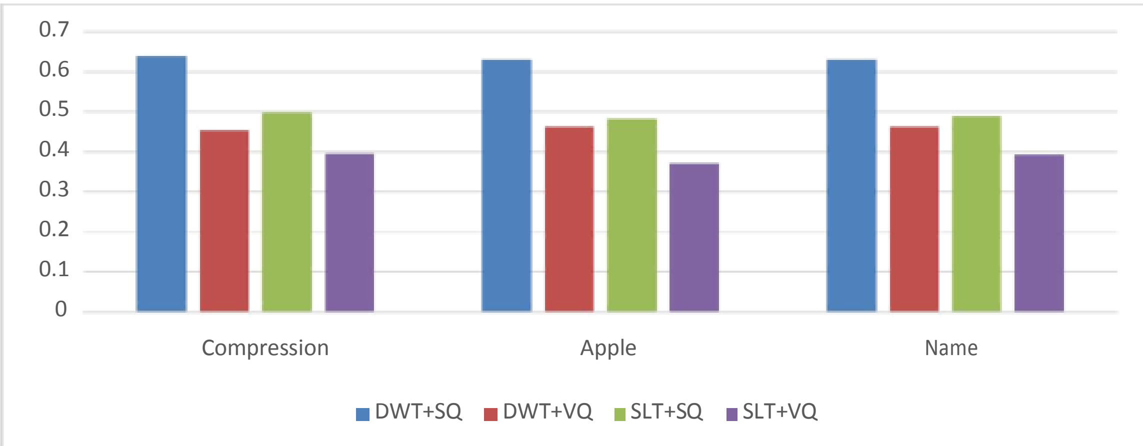

Fig. 1.7: Comparison of CR between DWT and SLT.

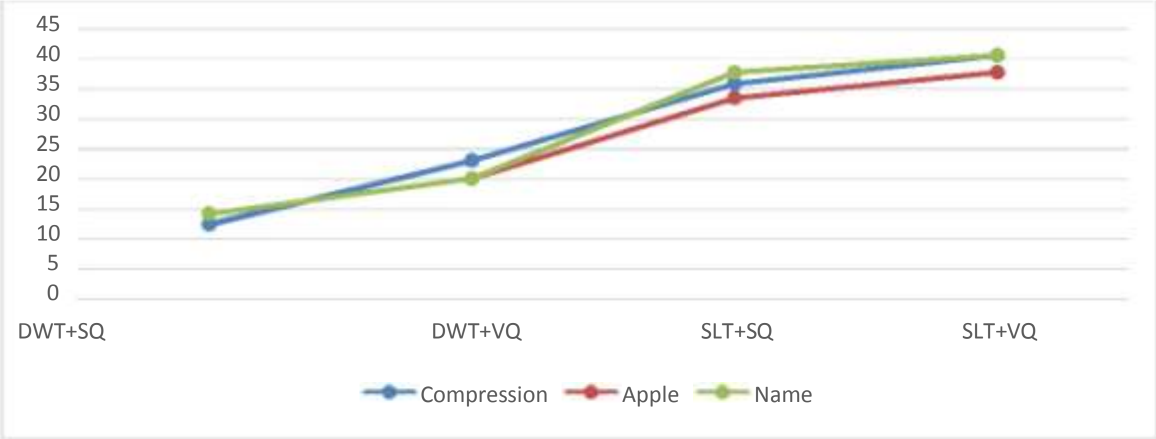

Fig. 1.8: Comparison of SNR between DWT and SLT.

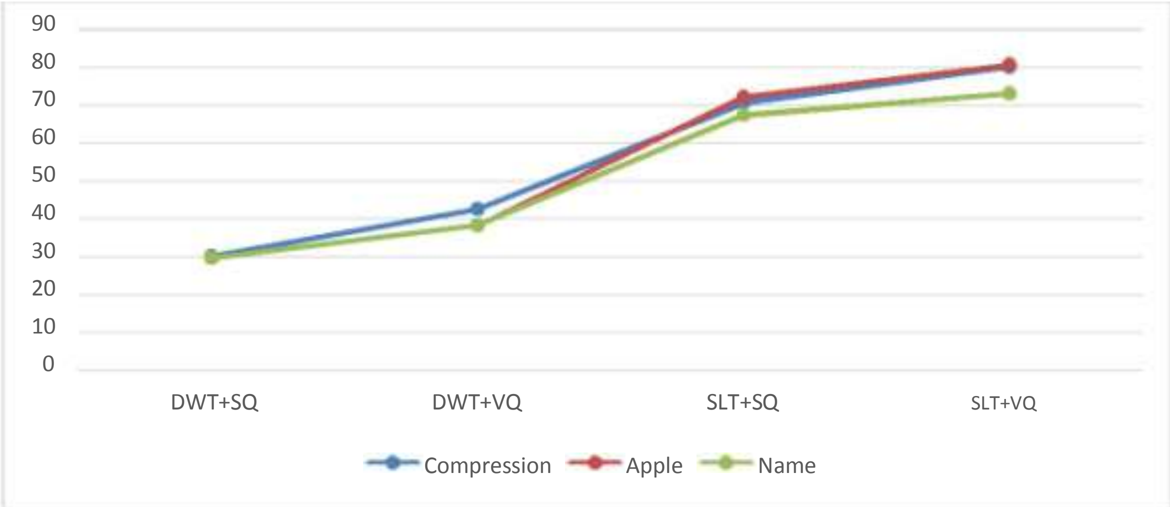

Fig. 1.9: Comparison of PSNR between DWT and SLT

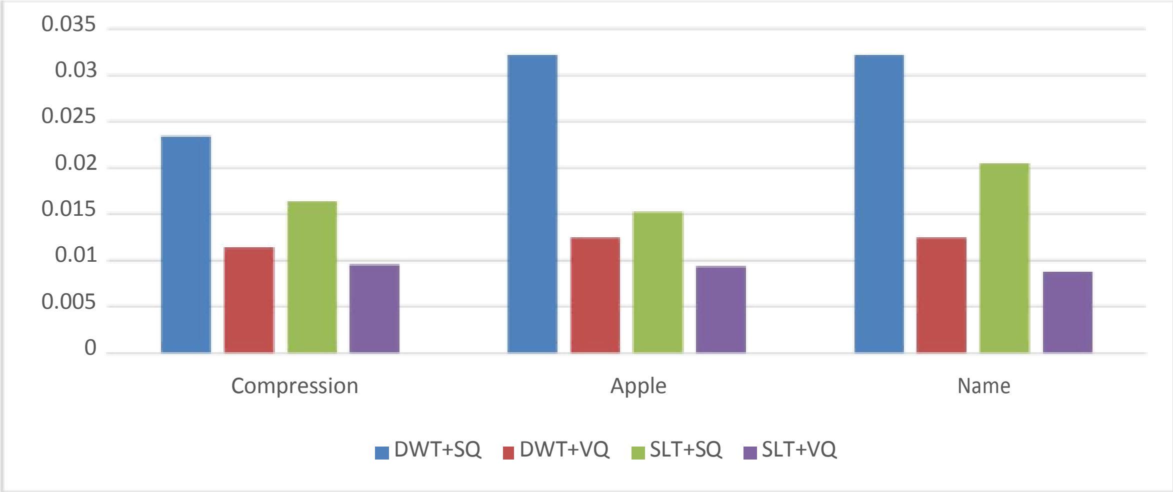

Fig. 1.10: Comparison of NRMSE between DWT and SLT.

The comparison between the proposed methodology and DWT method is listed in Table 1.3. Fig. (1.7-1.10) depicts the comparison between fidelity assessment parameters CR, SNR, PSNR and NRMSE respectively and it can be shown that the proposed method give better compression ratio and other fidelity parameters. Variations in these fidelity parameters with different threshold values are shown in Fig. (1.11-1.14). Thus, it is obvious that the developed method can be effectively used for speech compression.

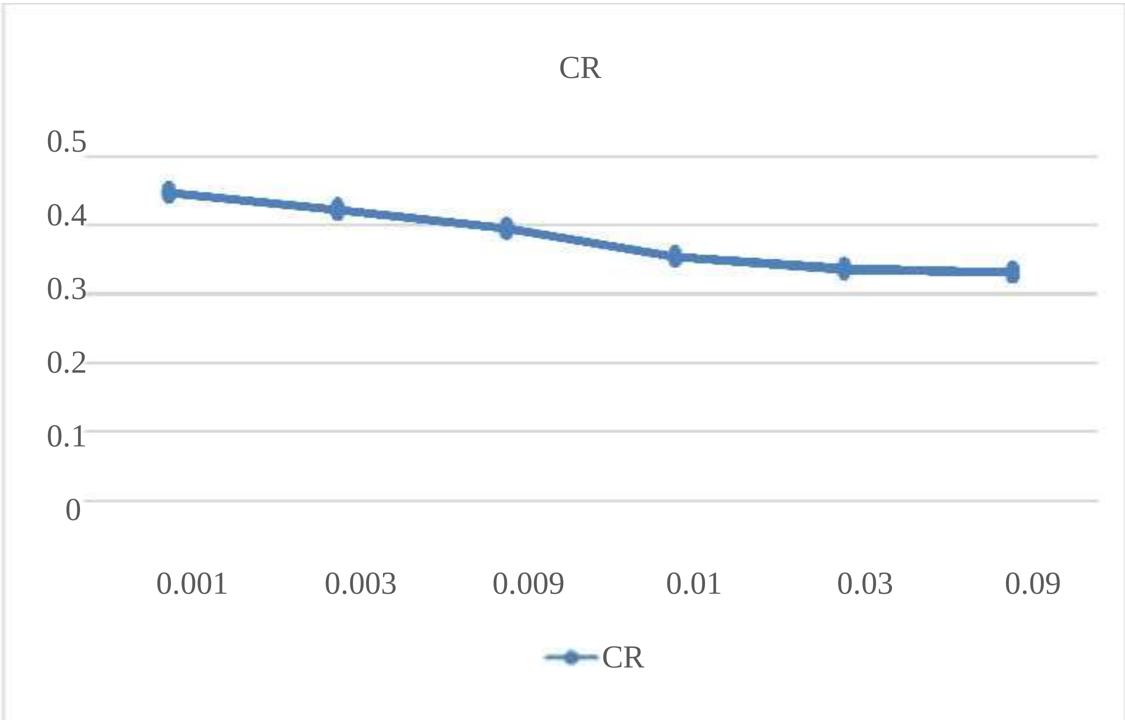

Fig. 1.11: Comparison of CR at different threshold values in proposed method.

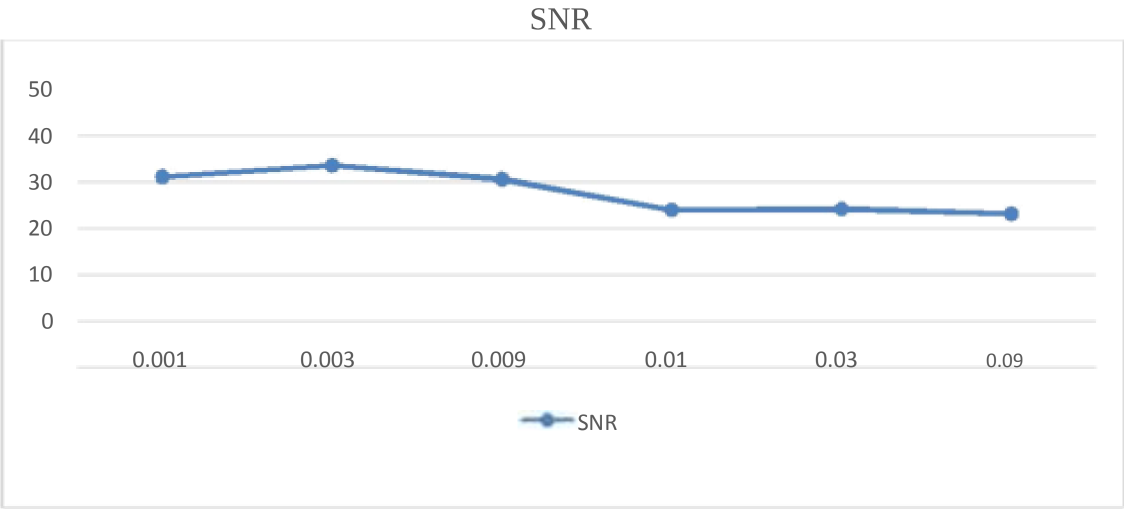

Fig. 1.12: Comparison of SNR at different threshold values in proposed method.

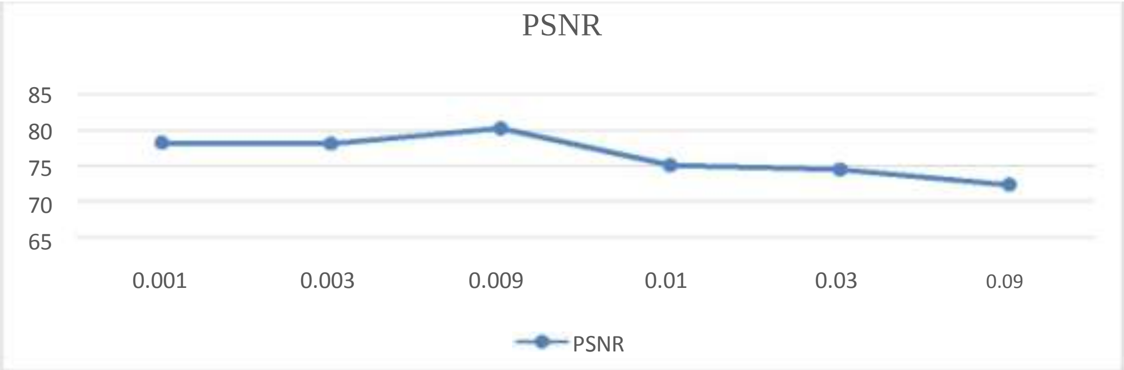

Fig.1.13: Comparison of PSNR at different threshold values in proposed method.

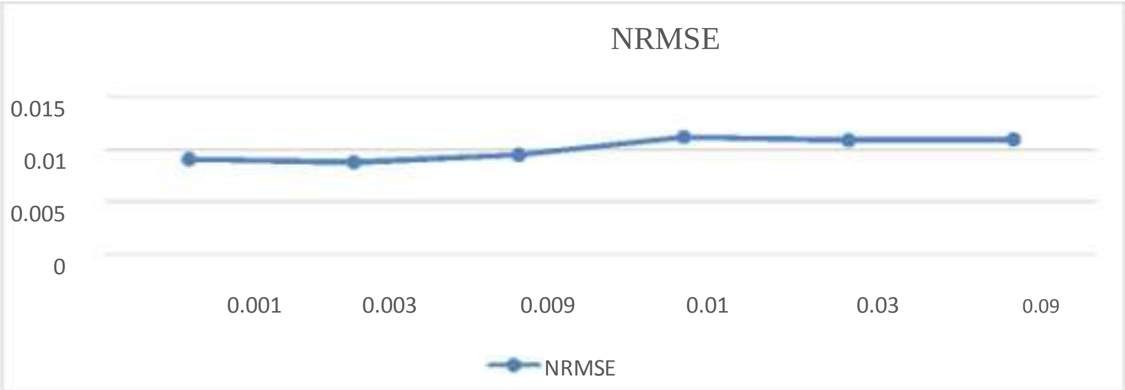

Fig. 1.14: Comparison of NRMSE at different threshold values in proposed method.

1.6 CONCLUSION:

In this chapter, a suitable method for speech compression based on the slantlet transform using different quantization and huffman encoding has developed. The simulation results included in this paper clearly show the key advantageous features of the developed method over others in the field of speech processing. It is also found that the developed method significantly improves the quality of reconstruction of the compressed speech signal, and also provides the comparable compression ratio as compared to other existing method. Therefore, it is concluded that it can be very effectively used for speech compression.

REFERENCES:

1. A.Gersho, “Advance in speech and audio compression” , Proceedings of the IEEE, Vol. 8, pp.900–918, 1994.
2. Overview of the human speech system,kom.aau.dk/group/04gr742/pdf/ speech_production.pdf
3. A.Gersho, “Speech coding”. In A. N. Ince (Ed.), Dig ital speech processing, Boston: Kluwer Academic, Vol. 4, pp.73–100, 1992.
4. N.Junejo, N.Ahmed, M. A.Unar and A.Q.K Rajput, “Spe ech and image compression using discrete wavelet transform”, IEEE symposium on adva nces in wired and wireless communication, Vol. 8, pp. 45-48, 2005.
5. H.M Magboun, N.Ali, M.A Osman and S.A Alfandi, “Mul timedia speech compression techniques”, IEEE international conference on compu ting science and information technology (ICCSIT), Vol. 9, pp. 498-502, 2010.
6. M.Arif, and R.S Anand, “Turning point algorithm for speech signal compression”, International journal of speech technology, doi: 10.1007/s10772-012-9151-7, 2012.
7. E.Shlomot, V.Cuperman, and A.Gersho, “Combined harm onic and waveform coding of speech at low bit rates” In IEEE conference on acou stics, speech and signal processing (ICASSP98), Vol. 2, pp. 585–588, 1998.
8. E.Shlomot, V.Cuperman, and A.Gersho, “Hybrid coding : combined harmonic and waveform coding of speech at 4 kb/s”, IEEE transactions on speech and audio processing, Vol 9(6), pp.632–645, 2001.
9. E.N.Zois and V. Anastassopoulos, “Morphological wav eform coding for writer identification”, Pattern recognition, Vol. 33(3), p p.385–398, 2000.
10. McCauley, J. Ming, J. Stewart and P. Hanna, “Subban d correlation and robust speech recognition”, IEEE transactions on speech and audio processing, Vol. 13(5), pp. 956-964, 2005.
11. K. Ramchandran, M. Vetterli, C. Herley, “Wavelet, s ubband coding and best bases”, Proceedings of the IEEE, Vol. 84(4), pp.541-560, 1996.
12. P.Vankateswaran, A.Sanyal, S.Das, R.Nandi, and S.K Sanyal, “An efficient time domain speech compression algorithm based on LPC and sub-band coding techniques”, Journal of communication, Vol. 4(6), pp. 423-428, 2009.
13. R.H Laskar, K.Banerjee, F. A Talukdar and K .Sreenivasa Rao, “A pitch synchronous approach to design voice conversion system using source-filter correlation”, International journal of speech technology, Vol.15, pp. 419-431, 2012.
14. M. A. Shahin, “Speaker identification investigat ion and analysis in unbiased and biased emotional talking environments”, International jour nal of speech technology, Vol.15, 325– 334, 2012.
15. G. Rajesh, A. Kumar and K. Ranjeet, “Speech compres sion using different transform techniques”, International conference on computer a nd communication technology (ICCCT), 2011.
16. Kumar, G.K. Singh, G. Rajesh and K. Ranjeet, “Th e optimized wavelet filters for speech compression”, International journal of speec h technology, 2012.
17. J. I. Agbinya, “Discrete wavelet transforms techniq ues in speech processing”, IEEE Tencon digital signal processing applications proceedings, pp. 514-519, New York, 1996.
18. Amara Graps, “An Introduction to wavelet”.
19. S.G Mallat, “A theory for multiresolution signal de composition: the wavelet representation”, IEEE Transactions on pattern analy sis and machine intelligence, Vol.11, pp. 674–693, 1989.
20. Kumar, A. K. Bhandari, and P.K. Padhy, “Improved Normalized Difference Vegetation Index Method Based on DCT and SVD for Satellite Image Processing”, IET Signal processing, Vol. 6, pp. 617-625, 2012.
21. Y. Yusong, W. Chunmei, Su Guangda, S. Qingyun “Inve rtible Integer DCT Applied on Progressive until Lossless Image Compression”, IEEE Proceeding of the 3rd International symposium on image and signal processing and analysis (ISPA 2003), vol. 2, pp. 1018-1023, 2004.
22. N. Al-Hinai, K. Neville, A. Z. Sadik and Z. M. Hussain, “Compressed Image Transmission over FFT-OFDM: A Comparative Study”, I EEE Proceeding on telecommunication networks and application conference (ATNAC), pp.465-469,