International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume: 3 | Issue: 3 | Mar-Apr 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 - 6470

Study on Speech Compression and Decompression by using Discrete Wavelet Transform

Sandar Oo

Associate Professor, Faculty of Natural Science, University of Computer Studies, Loikaw, Myanmar

How to cite this paper: Sandar Oo "Study on Speech Compression and Decompression by using Discrete Wavelet Transform" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-3, April 2019, pp.252-258, URL: http://www.ijtsrd.com/papers/ijtsrd21727.pdf

IJTSRD21727

ABSTRACT

Speech signal can be compressed and decompressed by discrete wavelet transform technique. Discrete wavelet transform compression is based on compressing speech signal by removing redundancies present in it. Speech compression is a technique to transform speech signal into compact form. Objective of compressing speech signal is to enhance transmission and storage capacity. The compression parameters in speech such as Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR), Normalized Root Mean Square Error (NRMSE), Compression Factor (CF) and Retained Signal Energy (RSE) are measured using Matlab.

Keywords: *Discrete wavelet transform, Denoising, Soft thresholding, Hard thresholding*

1. INTRODUCTION

Speech compression is the process of representing a voice signal for efficient transmission or storage. The compressed speech can be sent over both band limited wire and wireless channels. The aim of speech compression is to represent the samples of a speech signal in a compact form thus having the less code symbols without degrading the quality of the speech signal [1]. The compressed speech is very important in cellular and mobile communication. It is also applied in voice over internet protocol (VOIP), videoconferencing, electronic toys, archiving, digital simultaneous voice and data (DSVD), numerous computer-based gaming and multimedia applications [2].

Speech signal is compressed by converting the signal data into a new format that requires less bits to transmit. There are two basic categories of compression techniques. The first category is lossless compression. Lossless compression methods achieve completely error free decompression of the original signal. The second category is lossy compression. A lossy compression method produces inaccuracies in the decompressed signal. Lossy techniques are used when these inaccuracies are so small as to be imperceptible. The advantage of lossy technique over lossless one is that much higher compression ratios can be attained. With wavelet compression method, the imperceptible inaccuracies can be found in the decompressed signal [3].

Wavelet analysis has the benefit of varying the window size. This means that wavelets can efficiently trade time resolution for frequency resolution and vice versa. Wavelets can adapt to various time-scales and perform local analysis. Furthermore, wavelets have the ability to detect characteristics of non-stationary signals due to their finite nature that describes local features. Wavelets have been widely applied to areas such as speech and image denoising and compression [4,5]. Wavelet compression is a form of predictive compression where the amount of noise in the data set can be estimated relative to the predictive function [6].

Speech compression is the technology of converting human speech into an efficiently encoded representation that can later be decoded to produce a close approximation of the original signal. Figure 1 shows the block diagram used for compression of the speech signal and reconstruction of the signal.

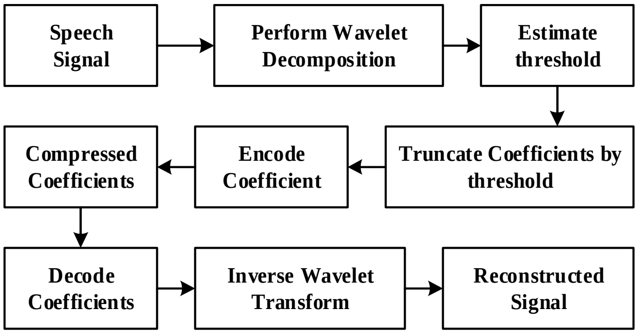

Figure 1 Block diagram used for compression and reconstruction of the speed signal.

Wavelet analysis is not a compression tool but a transformation to a domain that provides a different view of the data that is more suitable to compression than the original data itself. First the speech signal is decomposed into the wavelet transform coefficients. Then a threshold is calculated and applied to the wavelet coefficients. The small valued coefficients below a threshold are truncated to zero made an imperceptible to the signal. Signal compression is achieved by encoding the thresholder coefficients.

Many of the wavelet coefficients produced from the wavelet transform have an absolute value close to zero. These small valued coefficients are likely to attribute only small variations of the signal and contain a small percentage of the signal's total energy. These small coefficients can be discarded without a significant loss in the quality of the signal and more importantly of the interesting features. Thus, a threshold is required below which all coefficients will be discarded. The compressed signal is decoded. And then the decoded signal must be reconstructed by the inverse wavelet transform to get the original signal.

The rest of this paper is arranged as follows. In Section 2, speech compression using discrete wavelet transforms related literature to identify the key issues and summarize the experiences from various studies in different countries about the topic. In Section 3, we describe the data, the methodology and present related descriptive statistics. In Section 4, compression factors associated with fatigue driving and/or the severity of fatigue-related crashes are reported. Discussion of results is given in Section 5.

2. Speech Compression Using Discrete Wavelet Transform

Speech compression using wavelets is primarily linked to the relative scarceness of the wavelet domain representation for the signal. Wavelets concentrate speech information (energy and perception) into a few neighbouring coefficients. As a result of taking the wavelet transform of the signal, many coefficients will either be zero or have negligible magnitudes. Data compression is then achieved by treating small valued coefficients as insignificant data and discarding them. The choice of wavelet, decomposition level in the discrete wavelet transform, threshold criteria for the truncation of coefficients and encoding coefficients are investigated for the process of compressing speech signal.

In the wavelet transform compression, the signal can be transformed into a wavelet domain of the signal. All values of the transform coefficients which lie below some threshold value are set to zero. Only the significant, non-zero values of the transform coefficients can be transmitted. This should be a much smaller data set than the original signal. At the receiving end, the inverse wavelet transform of the transmitted data will be performed by assigning zero values to the insignificant values which were not transmitted. This decompression produces an approximation of the original signal [3,9]. The measurement of the compression parameters is evaluated in terms of Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR), Normalized Root Mean Square Error (NRMSE) and Compression Factor (CF). The source code for speech compression will be written by using Matlab.

3. Methodology

3.1. Performance Measurement of Speech Compression

A number of compression parameters can be used to evaluate the performance of the wavelet-based speech compression, in terms of both reconstructed signal quality after decoding and compression. The parameters are

- Signal to Noise Ratio (SNR)
- Peak Signal to Noise Ratio (PSNR)
- Normalized Root Mean Square Error (NRMSE)
- Retained Signal Energy (RSE)
- Compression Factor (CF)

Signal to Noise Ratio: This value gives the quality of reconstructed signal.

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

σ_x^2 is the mean square of the speech signal and σ_e^2 is the mean square difference between the original and reconstructed signal.

Peak Signal to Noise Ratio: $PSNR = 10log_{10}\frac{NX^2}{\|x-r\|^2}$

N is the length of the reconstructed signal, X is the maximum absolute square value of the signal x and $\|x-r\|^2$ is the energy of the difference between the original and reconstructed signal.

Normalized Root Mean Square Error:

$$NRMSE = \sqrt{\frac{(x(n)-r(n))^2}{(x(n)-\mu_x(n))^2}}$$

$x(n)$ is the speech signal, $r(n)$ is the reconstructed signal, and $\mu_x(n)$ is the mean of the speech signal.

Retained Signal Energy: This indicates the amount of energy retained in the compressed signal as a percentage of the energy of original signal.

$$RSE = \frac{100 * \|x(n)\|^2}{\|r(n)\|^2}$$

$\|x(n)\|$ is the norm of the original signal and $\|r(n)\|$ is the norm of the reconstructed signal. For one dimensional orthogonal wavelets the retained energy is equal to the L^2-norm recovery performance.

Compression Factor:

It is the ratio of the original signal to the compressed signal. The value of compression factor greater than 1 indicates compression and less than 1 indicates expansion. I referred to previous theory in my research work.

4. Analytical Results

The mother wavelet chosen to compress speech signal is important as some wavelets offer better reconstruction quality and different compression ratios than others. However, there is no wavelet that gives the best results for all kinds of signals. The test signal is "Great, now we've got time to party". The test signal 'voice38kz.wav' is formed by converting the MP3 file of audio into wav file by 'wavesurfer' software. The 'voice38kz.wav' has 25913 sampled data with sampling frequency 8kHz.

Selecting mother wavelet is related to the amount of energy a wavelet basis function can concentrate into the first level approximation coefficients. The signal energy retained in the first N/2 transform coefficients is shown in Table 1.

Table1. Signal energy retained in the first N/2 transform coefficients

Wavelet	Haar (db1)	db2	db4	db6	db8	db10
Ea	96.3282	98.7937	99.4076	99.5122	99.5728	99.5949
Ed	3.6718	1.2063	0.5924	0.4878	0.4272	0.4051

This energy is equivalent to the energy stored in the first level approximation coefficients. The higher the amount of energy in the first level approximation, the better is the wavelet for compression of the signal. The Haar and Daubechies (db2, db4, db6, db8, db10) wavelets concentrate more than 96% of the signal energy. Db10 wavelet concentrates 99.5949 % of energy into the first level approximation coefficients. Wavelets with many vanishing moments should be utilized for better reconstruction quality as less distortion and more signal energy concentration are introduced in the approximation coefficients. Wavelets with many vanishing moments are described with many coefficients in the scaling and wavelet functions. Thus, the computation of the wavelet transforms, the complexity of the algorithm and the output file size are increased. Figure 3 shows the flow chart of the program for compression of the speech signal.

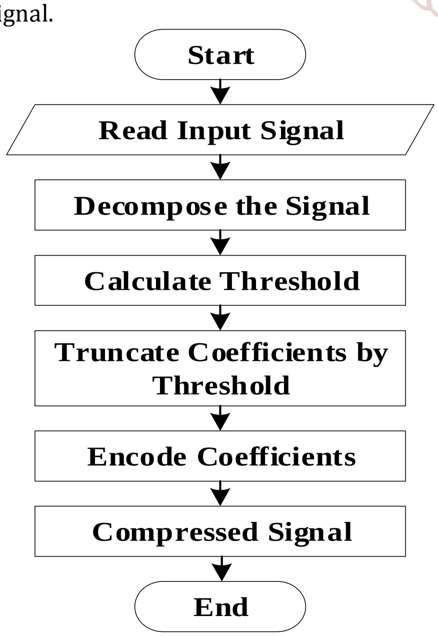

Figure3. Flow chart of the program for compression of the speech signal

In this work, the six wavelets are chosen and compared for speech compression. Choosing a decomposition level for the discrete wavelet transform usually depends on the type of signal being analyzed. For processing speech signal no advantage is gained in going beyond level 5[8]. After calculating the wavelet transform of the speech signal, compression involves truncating wavelet coefficients below a threshold. For the truncation of small valued transform coefficients, level dependent thresholding is used. Haar and Daubechies (db2, db4, db6, db8, db10) wavelets are used and compared against each other to measure the compression parameters for the speech signal. The signal is decomposed at scale 5 and level dependent threshold is applied. Figure 4 shows the flow chart of the program for decompression of the signal.

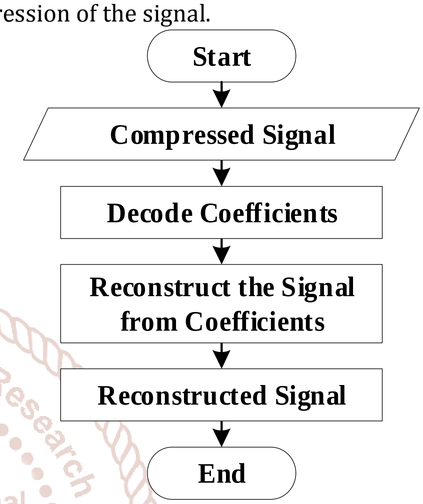

Figure4. Flow chart of the program for decompression of the speech signal

The flow chart of the program for calculation of compression parameters is shown in Figure 5.

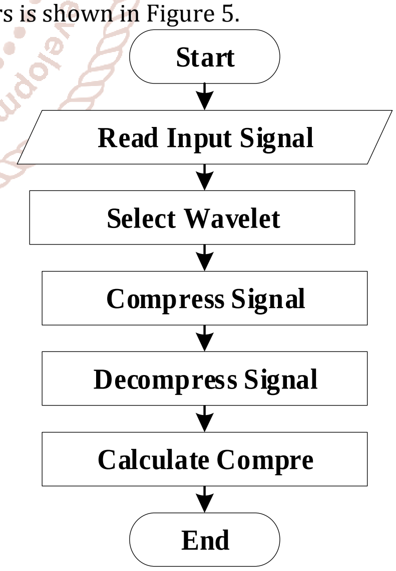

Figure5. Flow chart of the program to calculate the compression parameters

The results of the compression parameters are shown in Table 2 to Table 6. The quality of the reconstruction signal is compared with the original signal by using SNR, PSNR and NRMSE. The higher SNR, PSNR and the lower NRMSE values give the better quality of reconstructed signal. Db10 wavelet gives the higher value of SNR, PSNR and the lower value of NRMSE than other wavelets in decomposition level 1.

Table2. Measurement of compression parameters with different wavelets at level 1 decomposition

Wavelet	RSE (%)	Zeros (%)	SNR	PSNR	NRMSE	CF
Haar	99.9627	33.4337	34.2806	47.0935	0.0193	1.3229
db2	99.9916	33.4118	40.7657	53.5786	0.0092	1.3272
db4	99.9960	33.3951	43.9631	56.7760	0.0063	1.3331
db6	99.9971	33.3783	45.3533	58.1661	0.0054	1.3409
db8	99.9974	33.3655	45.8720	58.6848	0.0051	1.3412
db10	99.9977	33.3488	46.4274	59.2403	0.0048	1.3461

In decomposition level from 2 to 5, the SNR, PSNR values of db10 wavelet are not obviously higher than db8 wavelet. But db10 wavelet is better than db8. Thus, db10 wavelet gives the best result among other wavelets. RSE is the amount of energy retained in the compressed signal as a percentage of the energy of original signal. RSE is over 95% for decomposition level up to 3. The value of RSE is lesser in decomposing at scale 4 and the least value at scale 5. The compression factor and the % of zero coefficients are increased with increase in decomposition level. Figure 6 shows the original signal.

Table3. Measurement of compression parameters with different wavelets at level 2 decomposition

Wavelet	RSE (%)	Zeros (%)	SNR	PSNR	NRMSE	CF
Haar	99.0978	61.8985	20.4471	33.2600	0.0950	2.0556
db2	99.7497	61.8759	26.0150	38.8279	0.0500	2.0970
db4	99.8829	61.8453	29.3154	42.1283	0.0342	2.1634
db6	99.9072	61.8185	30.3266	43.1394	0.0305	2.1732
db8	99.9253	61.7917	31.2682	44.0811	0.0273	2.1772
db10	99.9315	61.7765	31.6415	44.4544	0.0262	2.1929

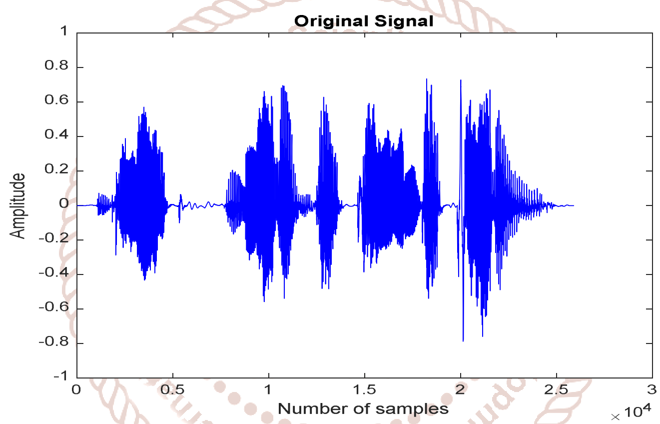

Figure6. The original signal

The comparison between the original signal and reconstructed signal using db10 wavelet is shown in Figure 7 for decomposition level 3, in Figure 8 for level 4 and in Figure 9 for level 5 respectively.

Table4. Measurement of compression parameters with different wavelets at level 3 decomposition

Wavelet	RSE (%)	Zeros (%)	SNR	PSNR	NRMSE	CF
Haar	95.4534	79.3718	13.4231	26.2360	0.2132	3.5037
db2	98.0735	79.3519	17.1524	29.9652	0.1388	3.6080
db4	98.9267	79.3198	19.6928	32.5056	0.1036	3.7934
db6	99.0711	79.3054	20.3203	33.1332	0.0964	3.8395
db8	99.2110	79.2888	21.0290	33.8419	0.0888	3.8659
db10	99.2258	79.2783	21.1116	33.9244	0.0880	3.8670

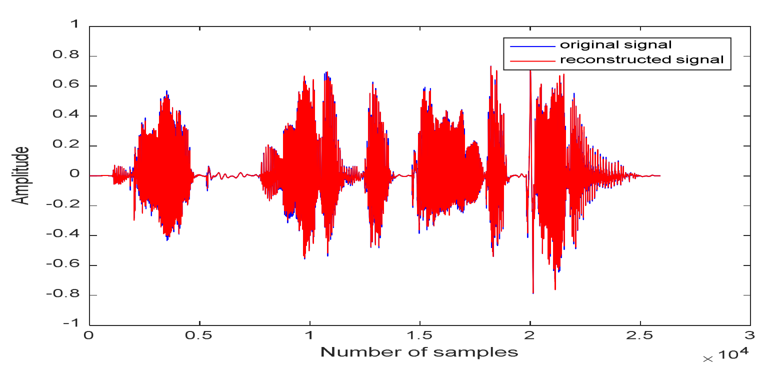

Figure7. The comparison between the original signal and reconstructed signal at decomposition level 3

Table5. Measurement of compression parameters with different wavelets at level 4 decomposition

Wavelet	RSE (%)	Zeros (%)	SNR	PSNR	NRMSE	CF
Haar	85.3250	89.1187	8.3342	21.1471	0.3831	6.2957
db2	89.4580	89.0985	9.8080	22.6209	0.3233	6.4557
db4	91.3590	89.0825	10.6344	23.4472	0.2940	6.7412
db6	91.7306	89.0576	10.8252	23.6381	0.2876	6.8336
db8	91.9858	89.0455	10.9614	23.7743	0.2831	6.7210
db10	92.1907	89.0137	11.0739	23.8867	0.2795	6.7888

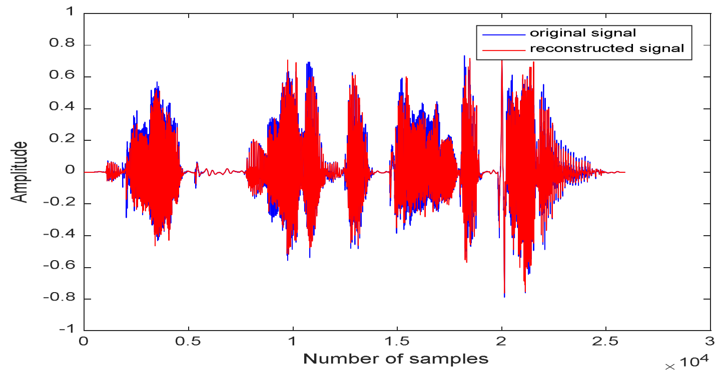

Figure8. The comparison between the original signal and reconstructed signal at decomposition level 4

Table6. Measurement of compression parameters with different wavelets at level 5 decomposition

Wavelet	RSE (%)	Zeros (%)	SNR	PSNR	NRMSE	CF
Haar	65.7828	94.3471	4.6576	17.4704	0.5850	11.5631
db2	70.7420	94.3259	5.3375	18.1504	0.5409	11.5528
db4	72.6046	94.3036	5.6232	18.4361	0.5234	11.6936
db6	73.5452	94.2692	5.7750	18.5878	0.5143	11.8541
db8	74.3177	94.2392	5.8037	18.6165	0.5068	11.6883
db10	74.2488	94.2167	5.8920	18.7049	0.5075	11.5169

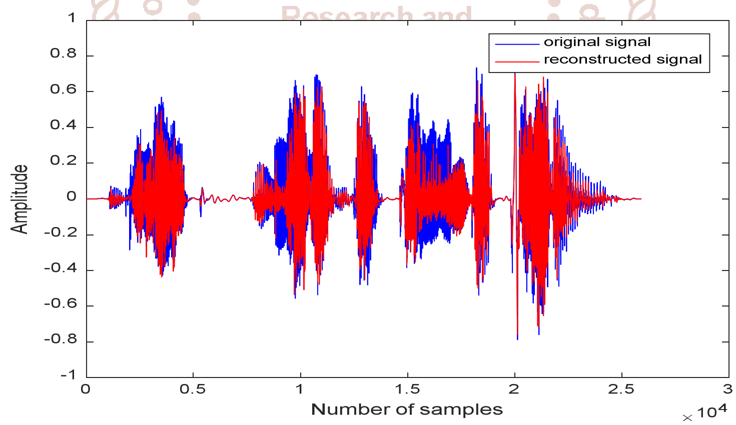

Figure9. The comparison between the original signal and reconstructed signal at decomposition level 5

5. Results and Discussion

The reconstructed signal is written to an audio file by using 'audiowrite' function in Matlab. Listening test is carried out on each level of the reconstructed signal in audio file. The quality of reconstructed signal is very close to the original signal in the decomposition level 1 and 2. The quality of the signal is nearly close to the original signal in level 3. The quality of the reconstructed signal is bad at decomposition level 4 and 5. From the overall results the level 3 decomposition is suitable for this signal. At higher levels the approximation data is not as significant and hence does a poor job in approximating the input signal. The number of samples in the compressed signal with different wavelets are shown in Table 7.

Table 7. The number of samples in the compressed signal by six wavelets for decomposition level up to 5

Wavelet	Level 1	Level 2	Level 3	Level 4	Level 5
Haar	19588	12606	7396	4116	2241
db2	19525	12357	7182	4014	2243
db4	19438	11978	6831	3844	2216
db6	19325	11924	6749	3792	2186
db8	19321	11902	6696	3799	2217
db10	19250	11817	6701	3817	2250

Figure 10 shows the results of compression parameters using db10 wavelet with different decomposition levels.

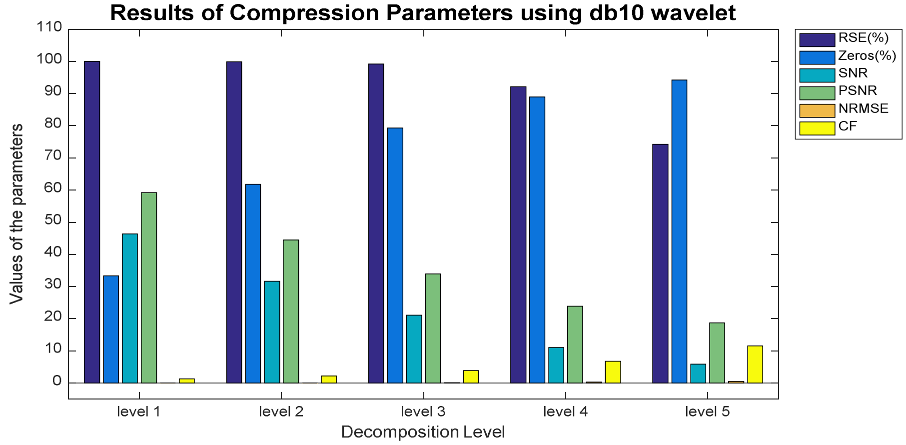

Figure10. The results of compression parameters using db10 wavelet with different decomposition levels

The comparison result of the original signal and compressed signal is shown in Figure 11.

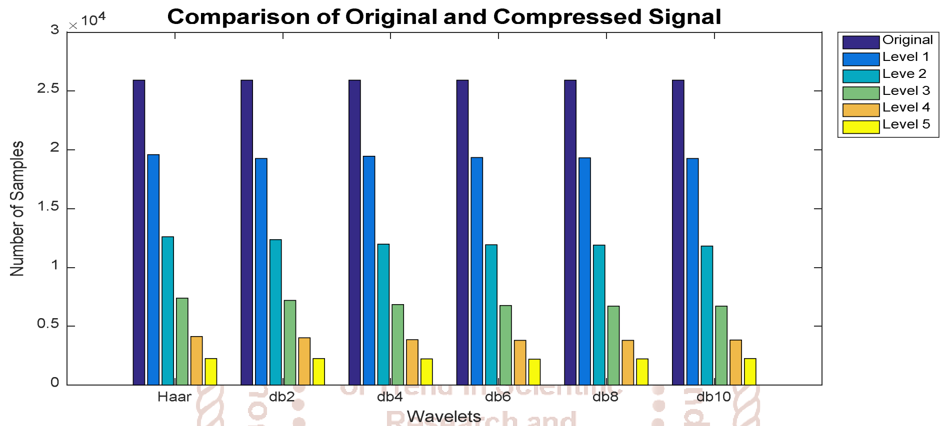

Figure11. The comparison of the original and compressed signal at decomposition level up to 5

The signal to noise ratio variation relative to compression factors using db10 wavelet is shown in Figure 12. The source code for the calculation of compression parameters is displayed in Matlab.

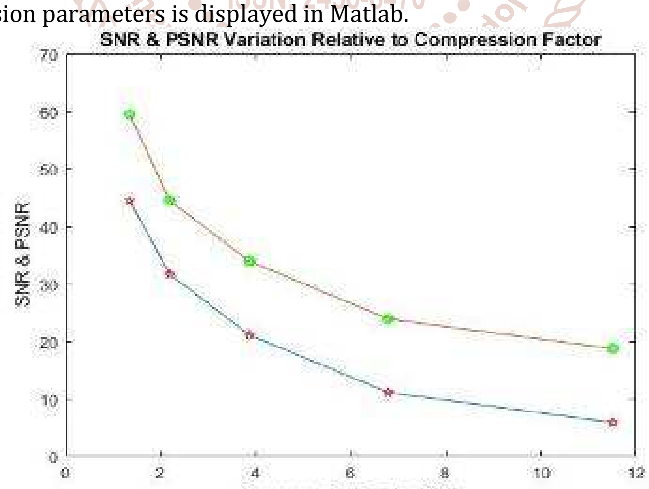

Figure 12.SNR and PSNR variation relative to compression factors using db10 wavelet

6. Conclusion

Speech compression is a solution to the problem of large amount of storage and bandlimited transmission. The discrete wavelet transform performs well in the compression of speech signal. The performance measurement results are obtained by using the Haar and Daubechies wavelets. The compressed signal can be reconstructed back to its original form with full audibility. A good reconstructed signal is the one with low MSE and high PSNR and SNR. This means that the signal has low error and high signal fidelity. Db10 wavelet has the high SNR , PSNR values and the low NRMSE as compared with other wavelets. SNR, PSNR, NRMSE, CF, RSE and % of zero coefficients are measured to evaluate the performance of the speech compression. Decomposition level at scale 3 is suitable for this signal. The measurement results are obtained by writing the source code in Matlab.The decomposition level for different types of speech will be chosen using wavelets.

7. References

[1] Jagtap S.K., Mulye M S., and Uplane M.D., (2015), " Speech Coding Techniques", the 4th International Conference on Advances in Computing, Communication and Control, London: Elsevier Inc./S&T Books and Cell Press

[2] Sun L., Mkwawa I.H., Jammeh E., and Ifeachor E., (2013), "Guide to Voice and Video over IP", Computer Communications and Networks, UK: Springer-Verlag London Ltd.

[3] WALKER J.S., (2008), "A Primer on WAVELETS and their Scientific Applications", 2nd Edi, USA: Chapman & Hall/CRC Press

[4] Kyriakopoulos, K.G., and Parish D.J., (2010), " Applying Wavelets for the Controlled Compression of Communication Network Measurements", Loughborough University's Institutional Repository, pp. 507-520

[5] Michel M., Yves M., Georges O., and Poggi J.M.,(2010), "Wavelet Toolbox User's Guide", the Math Works, Inc.

[6] Petkar H, (2017), "Speech Compressing using Wavelet Transform", IOSR Journal of Computer Engineering, Vol. 19, Issue 3, May, pp.33-41

[7] Rajput A,Q., "Wavelet Based Speech Compression for VOIP Applications", Institute of Information Technology, Mehran University of Engineering & Technology, Jamshoro. Sindh Pakistan

[8] Rao N., and Homer J., (2001), " Speech Compression using Wavelets", Information Technology and Electrical Engineering, the University of Queensland, October

[9] Winkler J.R., "Orthogonal Wavelets via Filter Banks: Theory and Applications", the University of Sheffield, Department of Computer Science, United Kingdom, pp.1-22.