International Journal of Electronics Communication and Computer Engineering
Volume 3, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Speech Compression and Enhancement using Wavelet Coders

Satish Kumar
Deptt. of ECE
Amity School of
Eng. & Technology
Amity University U.P.
Lucknow Campus, India
skumar2@lko.amity.edu

O. P. Singh
Deptt. of ECE
Amity School of
Eng. & Technology
Amity University U. P.
Lucknow Campus, India
opsingh@amity.edu

G. R. Mishra
Deptt. of ECE
Amity School of
Eng. & Technology
Amity University U. P.
Lucknow Campus, India
gr_mishra@rediffmail.com

Saurabh Kumar Mishra
Deptt. of ECE
Amity School of
Eng. & Technology
Amity University U. P.
Lucknow Campus, India
saurabhmishra18@gmail.com

Akanksha Trivedi
Deptt. of ECE
Amity School of
Eng. & Technology
Amity University U. P.
Lucknow Campus, India
akanksha2trivedi@gmail.com

Abstract - **In the field of communication system there is always a trade-off between compression ratio (CR) and quality of signal when it comes to the lossy compression technique. In digital telephony system speech compression and high-quality of speech is crucial, because with a very limited bandwidth, service is provided with a huge number of users, so to improve the efficiency of bandwidth and for higher SNR an effective speech coder is required. In recent years discrete wavelet transform (DWT) emerged as a very effective technique for image and speech signal analysis. This paper presents wavelet compression technique by which the ratio of compression and signal to noise of a speech signal can be balanced. The results have been simulated on Mat Lab toolbox 7.10, clearly displays the values of SNR and CR for male and female speech signals using Haar and db6 wavelet.**

Keywords - **Signal to Noise Ratio, Discrete Wavelet Transform, Wavelet coefficient and Compression Ratio.**

I. Introduction

Speech coding has always been and still a major issue in the terms of maintaining a ratio between signal to noise and compression. Discrete wavelet transform has been used successfully for image compression but a very low attention has been paid for implementing this technique in the field of speech coding. In telecommunication field, speech compression is an issue directly related to the service provider but not a direct problem of users, users want higher SNR. Linear predictive coding is mostly used for speech coding that yields higher compression ratio (fixed) but by using wavelet technique we get higher SNR than LPC and a variation in compression ratio. Speech compression using discrete wavelet transform [1] will be very useful in military applications, digital cellular telephony system because there is limited bandwidth (lower bit rate is required) and higher speech quality is essential, especially in military purpose. This technique is very much useful in video-conferencing [9] where a wide bandwidth is already occupied by video processing itself. There are many other applications like storage of speech signal and transmission of voice at a later time.

Speech coders may be classified into two categories [6]: waveform coding technique and source coding technique (analysis-synthesis coders). Waveform coding technique tries to copy the exact shape of electrical signal produced by microphone. Analysis-synthesis coders use an entirely different approach and no approach is made for producing exact shape at the receiver. PCM is a popular waveform coding technique while LPC [1] is for voice coding. In this paper a different technique is used for speech compression named by discrete wavelet transform. The discrete wavelet transform of a signal produces coefficients and we apply all data operations on coefficients according to our need.

II. Discrete Wavelet Transform

Wavelet [1] is a new technique for analyzing and compressing a speech signal, it is more advantageous technique because it holds both time and frequency aspect of the signal and have localize analysis of a larger signal. The basic concept behind wavelet is to analyze a signal according to the scale. The first essential thing is to choose a mother wavelet then any signal can be represented by its translated and scaled version. Wavelet breaks speech signal into few coefficients, so when we take wavelet transform of it, some of the coefficients becomes zero or have very small value. Data compression is done by treating small valued coefficients [12] insignificant and discarding them.

Discrete wavelet transform breaks the signal into high frequency and low frequency components. The output of high pass filter is known as detail coefficients and the output of low pass filter is approximation coefficients. Approximation coefficients are high scaled low frequency components while detail coefficients are low scaled high frequency components.

The definition of detail at level j

$$D_J(t) = \sum_{k\varepsilon Z} C(jk)\Psi_{j,k}(t)$$

The approximation at level j

$$A_J = \sum_{k>J} D$$

We distinguish the wavelet function Ψ from its scaling function Φ

and

$$\left(\int \psi(x)dx = 0\right)$$

$$\left(\int \phi(x)dx = 1\right)$$

Ψ is used to define details.

Φ is used to define approximations.

Furthermore, we can see that it gives another aspect of data than those obtained by traditional techniques. This technique can compress and de-noised signal without significant loss of data.

In order to compress the speech signal with higher SNR, we have adopted following significant steps represented with a block diagram in figure 1

International Journal of Electronics Communication and Computer Engineering
Volume 3, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

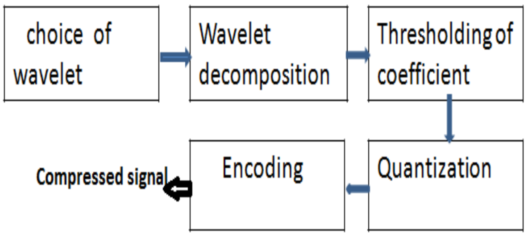

Fig.1. Steps involved in signal compression

III. Choice of Wavelet

The choice of wavelet is very important in wavelet compression technique, when it is used to design high quality speech coder. We cannot consider any wavelet better than other one because every wavelet is unique and has its own significance. Different kind of wavelet make trade-offs between the compact model of basis functions in space and how smooth they are. Mostly wavelets are classified by its vanishing moment and number of the vanishing moments is important for wavelet speech compressor. It is not practically implemented for real time applications with high number of vanishing moments [9] because the computational complexity of discrete wavelet transform is proportional to the number of vanishing moments.

With speech compression our objective is to improve SNR also, so wavelet can be selected on the basis of energy conservation properties in approximation coefficients. By using Daubechies D20, D12, D10 or D8 wavelets, 96% of the signal energy, level 1 approximation coefficients contains. The even numbers with D displays the number of coefficients and vanishing moment that are half of the number of coefficients. In this paper we have tested speech signal with Haar wavelet and db6 which contains 6 coefficients and 3 vanishing moments.

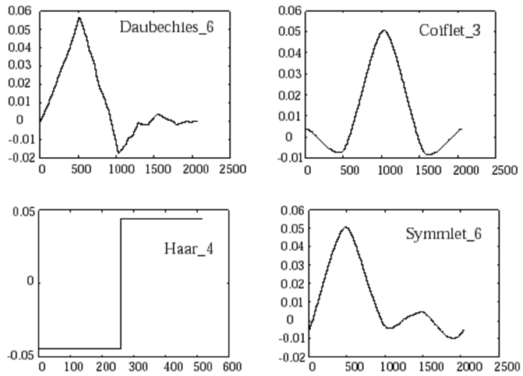

Fig.2. Different wavelet families [9]

IV. Wavelet Decomposition And Its Level

Unlike conventional techniques, wavelet decomposition breaks the input signal into high and low frequency components and the output is known as detail and approximation coefficients respectively. The output of low pass filter is further decomposed into high and low frequency components. This process is repeated to n levels; finally we have (n+1) outputs. Signal compression is done by selecting the approximation coefficients at the appropriate chosen level and most of the detail coefficients have non-essential information but very few coefficients have accurate signal component.

Choosing a decomposition level is depend upon the type of signal is to be analyzed and some other criterion such as entropy. If the frame based input is applied, then frame size must be a multiple of 2^n, where n represents the decomposition level.

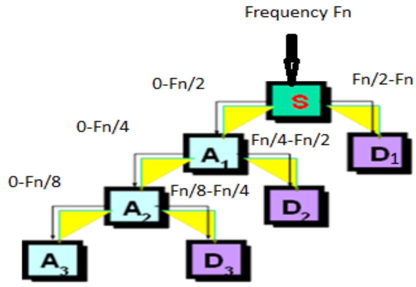

Fig.3. Wavelet decomposition

V. Thresholding of Coefficients

When the signal get decomposed into approximation and detail coefficients, then we truncate the small valued coefficients considering it to non- essential part of the signal. For truncating the coefficients there are two types of thresholding- hard and soft thresholding. Hard thresholding is used to compress the signal while soft thresholding is used to de-noise the speech signal. A Mat lab function *wdencmp* is used to wavelet coefficients thresholding for both compression and de-noising.

wdencmp enables us to choose whether it is global thresholding or level-dependent thresholding. Level-dependent thresholds are calculated using Brige-Massart strategy [3]. According to this strategy all the approximation coefficients are kept at the level of decomposition j. The numbers of detail coefficients that are to be kept at level i from 1 to j are given by the formula.

$$n_j = \frac{M}{(J+2-j)^a}$$

Typically, a = 1.5 for compression and a = 3 for de-noising.

If wavelet coefficients are high scarcely distributed then value of M is equal to L, where L is the length of coarsest approximation coefficients.

VI. Uniform Quantization

Quantization [7] is further processing of signal towards true compression. Quantization is to be done for reducing the information content in wavelet coefficients in such a manner that no error formed. In quantization process the step size depends on

1. Maximum value M_{max} in the signal matrix
2. Minimum value M_{min} in the signal matrix

International Journal of Electronics Communication and Computer Engineering
Volume 3, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

3. Number of quantization level L

When these parameters are available, then step size can be found

$$\Delta = \frac{M_{\max} - M_{\min}}{L}$$

Now the input is divided into L+1 levels with same interval size ranging from minimum value to maximum value to obtain quantization table. After completed quantization process, the quantized value will be fed into the next step of compression.

VII. Huffman Encoding

In quantization process, the quantized data contains some repeated data and that would be wastage of memory. Huffman coding is done to overcome this problem [5]. After Huffman coding, we get true compressed signal.

VIII. Performance Analysis

A. Wavelet coefficients of speech signal

When we performed DWT of signal using ‘db6’ and ‘Haar’ wavelet at level 6, we got 6-level approximation and detail coefficients. Decomposition of signal has been performed using Mat Lab toolbox 7.10 of a male spoken sentence ‘*compression of speech signal….*’ Using command

[C, L] = wavedec (s, 6,'db6');

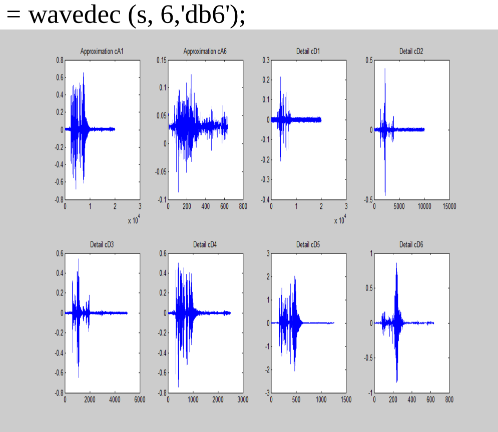

Fig.4. Wavelet coefficients using ‘db6’ at different level

Original speech signal s=cA1+cD1 where cA1 contains low frequency components and cD1 contains high frequency component of the original signal. When we listen cA1 using command **sound** (cA1, Fs), we are able to understand the voice with a low loss in quality.cD1 contains less essential information, in terms of coefficients.

B. Crude de-noising and Compression of a signal

De-noising [10] and compression of the speech signal has been performed on Mat Lab toolbox 7.10 of a male spoken sentence *“Compression of speech signal”.*

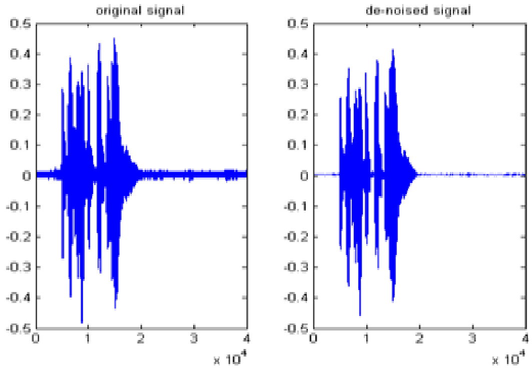

Fig.5. Original and de-noised signal using ‘db6’

Compression of the same signal (Spoken by male), has been performed by level dependent thresholding, using Brige-Massart strategy [3].

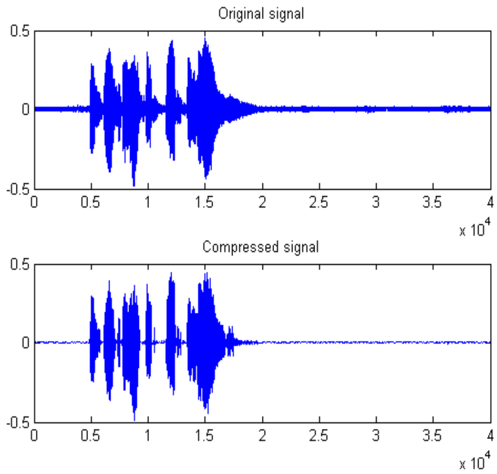

Fig.6. compressed signal for a male voice of 4 sec using ‘db6’ wavelet

This is another speech **(**spoken by female**)** hat has been compressed by the same wavelet technique.The female spoken sentence is *“Compression of speech signal”*.

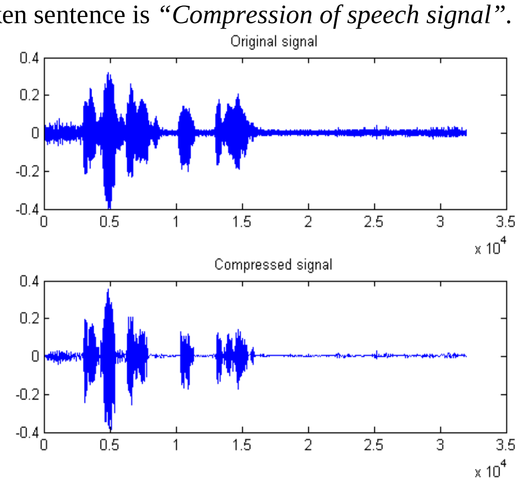

Fig.7. Compressed signal for a female voice of 3 sec using ‘db6’ wavelet.

International Journal of Electronics Communication and Computer Engineering
Volume 3, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

IX. Results

A. Compression ratio

Compression ratio is defined as the length of the original signal x (n) to the reconstructed signal r (n).

CR=Length x (n)/Length r (n).

B. Signal to noise ratio

$$SNR= 10 \log10 (\sigma_x^2 /\sigma_e^2)$$

Where 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 signals. Signal to noise ratio displays the quality of the speech signal. The greater the value of SNR, better the quality of speech signal**.**

Signal to noise ratio displays the quality of the speech signal. The greater the value of SNR, better the quality of speech signal**.**

Performance of male speech signal

Wavelet	Level	SNR	CR
Haar	6	12.0572	4.1538
Db6	6	16.6755	4.0187

Table I

Performance of female speech signal

Wavelet	Level	SNR	CR
Haar	6	11.7559	4.1820
Db6	6	15.3818	4.0759

Table II

We have tested on two speech signals, spoken by male and female respectively with two different wavelets 'Haar' and 'db6'. The results of compression are shown in table 1 and 2 in terms of CR and SNR.

We know that SNR and CR (in case of lossy compression) always have an inversely proportional relationship with each other, and that is why we wanted a balanced speech compression technique. With speech signal compression an enhanced speech quality is very much required. The most attractive feature of wavelet compression technique is that it has a variation in compression ratio so we can vary CR and SNR according to our need while mostly used technique like GSM coding technique has fixed compression ratio (1:5).

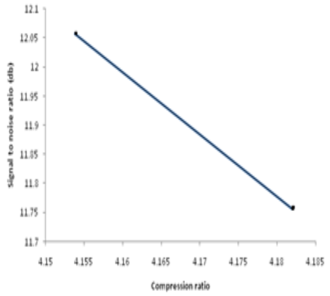

Fig.8. A graphical representation of performance of 'Haar' wavelet

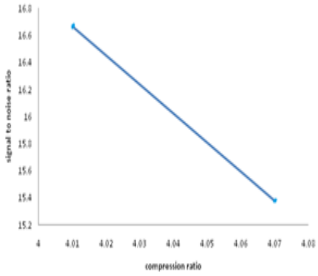

Fig.9. A graphical representation of performance of 'db6' wavelet

X. Conclusion

Compression of speech signal using wavelets enables us to vary the compression ratio as per our requirement. This paper justifies the inversely proportional relation between CR and SNR and applied discrete wavelet transform technique to maintain that trade-off. In this paper we used 'Haar' and 'db6' wavelets to decompose the signal at level 6 of two speech signals spoken by a male and a female. The compressed and de-noised signals shown upper clearly displays the difference with original speech signal while compression and de-noising performances of wavelet technique is shown in resulted table.

References

[1] Audio signal processing and coding, by Andreas Spanias, Ted Painter, and Venkatraman Atti

[2] http://www.data-compression.com/index.html

[3] Birgé, L., P. Massart "From model selection to adaptive estimation," in D. Pollard (ed.), *Festschrift for L. Le Cam*, Springer, (1997), pp. 55-88.

[4] Ben Gold and Nelson Morgan. 'Speech and Audio Signal Processing'. John Wiley and Sons, 2000.

[5] Abdul Mawla M. A. Najih, Abdul Rahman bin Ramli, V. Prakash, and Syed A.R,Speech Compression Using Discreet Wavelet Transform**,** 0-7803-7773-7/03/$17.00 *0* 2003 IEEE.

[6] Dr.V.Radha, Vimala.C, M.Krishnaveni**,** Comparative Analysis of Compression Techniques for Tamil Speech Datasets**,** 978-1-4577-0590-8/11/$26.00 ©2011 IEEE.

[7] AbduI Mawla M, A. Najih, Abdul Rahman Ramli, Azizah Ibrahim, Syed A.R, "Comparing Speech Compression Using Wavelets with Other Speech Compression Schemes", 0-7803-8 173-4/03/$17.0002003 IEEE.

[8] Dr. Hussein M. Magboub, Dr.Nasser, Ali, Dr.Mahmoud.A.Osman, Dr.S.A.Alfandi, "Multimedia Speech compression Techniques", 978-1-4244-5540-9/10/$26.00 ©2010 IEEE.

[9] Nikhil Rao, "Speech Compression Using Wavelets", ELEC 4801 Thesis Project, School of Information Technology and Electrical Engineering, Qld 4108, October 18, 2001.

[10] Donoho, D.L. (1995), "De-noising by soft-thresholding," *IEEE Trans. on Inf. Theory*, 41, 3, pp. 613-627.

[11] El-Bahlul Fgee, W. J. Phillips and W. Robertson, "comparing audio compression using wavelets with other audio compression schemes", *0-7803-5579-2/99/%10.00 c6 1999* IEEE.

[12] W. Kinsner and **A.** Langi, "Speech and Image Signal Compression with Wavelets", IEBE Wescanex Conference Proceedings, IEEE, New York, NY, pp. 368-375, 1993.

[13] R.V. *Cox* and P. Kroon, "Low bit-rate speech coders for multimedia communication", IEEE Communications Magazine, pages 34-40, 1996. http://www.bell-labs.com

[14] Cohen, A., I. Daubechies, J.C. Feauveau (1992), "Biorthogonal basis of compactly supported wavelets," *Comm. Pure Appli. Math.* , vol. 45, pp. 485-560.

AUTHOR'S PROFILE

Prof. Satish Kumar
M. Tech., FIETE is an alumnus and gold medalist of IIT Roorkee (Erstwhile University of Roorkee). He has worked as General Manager and Chief Vigilance Officer in India's first PSU, ITI Ltd. Bangalore. He has more than 34 years of industrial experience on state-of-the-art telecom equipment and infrastructure dealing with turn-key projects. His professional experience covers the whole spectrum of switching, transmission, access, telemetry and subscriber premises equipment. He also worked for providing design, system integration and Value Engineering solutions for various network based turn-key projects like Army Static Communication Network (ASCON), Integrated Electronic Warfare System (IEWS), Defence Communication Network (DCN), Tactical Communication System (TCS), Network For Spectrum (NFS) etc., Presently he is working as Professor, (Electronics & Communication Engineering) Amity University Lucknow campus. Email Id- skumar2@lko.amity.edu

Dr. O. P. Singh
obtained his Ph. D in Electronics Engineering for IT, BHU. He has over sixteen years experience and presently working as Professor and Head Department of Electronics & Communication Engineering, Amity School of Engineering & Technology, Amity University Uttar Pradesh Lucknow Campus. He is life member of Indian Society of Remote Sensing (ISRS) and Material Research Society of India (MRSI) . He has presented several papers in National Seminar/Conferences and also published papers in National/international Journals. His area of interest includes Microwave Remote Sensing Digital System, Signal Processing.
Email Id - opsingh@amity.edu

Dr. Ganga Ram Mishra
obtained his M.Sc. Electronics degree from Gorakhpur University and Ph. D in Electronics from Avadh University. He has over eleven years experience and presently working as faculty members in Department of Electronics & Communication Engineering, Amity School of Engineering & Technology, Amity University Uttar Pradesh Lucknow Campus. He is life member of Institution of Electronics & Telecommunication Engineers (IETE) and Material Research Society of India (MRSI).He has presented several papers in National Seminar/Conferences and also published papers in National/international Journals. His area of interest includes Digital System Design, Microprocessor Based System Design, Signal Processing, Embedded Systems and VLSI Design.
Email Id - gr_mishra@rediffmail.com

Mr. Saurabh Kumar Mishra
is currently pursuing M. Tech (EC&E) from Amity University Uttar Pradesh, Lucknow Campus, India. He has completed his graduation in 2010 from Uttar Pradesh Technical University.
His area of interest includes Speech Signal Processing, Image signal Processing. He has been actively involved in research works.
Emailid- saurabhmishra18@gmail.com

Ms. Akanksha Trivedi
is currently pursuing M. Tech (EC&E) from Amity University Uttar Pradesh, Lucknow Campus, and India. She has completed his graduation in 2011 from Uttar Pradesh Technical University. Her area of interest includes Speech Signal Processing, Image signal Processing & Digital Communication. She has been actively involved in research works.
Email Id- akanksha2trivedi@gmail.com