Australian Journal of Basic and Applied Sciences, 9(16) Special 2015, Pages: 229-234

ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

A Comprehensive Survey on Speech Compression Techniques Based on Discrete Orthogonal Transforms

[1]N. Vini Antony Grace, [2]Dr. A. Chilambuchelvan, [3]Dr. N. Usha Bhanu

[1] *Part-Time Research Scholar-Anna University, Chennai*
[2]*Professor-Department of CSE, RMK Engineering College, Kavaraipettai, Chennai-601206*
[3]*Associate Professor-Department of ECE, Saveetha school of Engineering, Saveetha University. Chennai - 600124*

ARTICLE INFO

Article history:
Received 20 January 2015
Accepted 02 April 2015
Published 20 May 2015

Keywords:
Speech compression,
Discrete Cosine Transform,
Discrete Wavelet Transform,
Discrete Walsh Hadamard Transform,
Wavelet Packet Decomposition.

ABSTRACT

Speech compression plays a very important role in various application fields such as mobile satellite communications, cellular telephony and teleconferencing. It is one of the leading vicinity of digital signal processing that spotlight on dipping the bit rate of speech signals for transmission and storage devoid of considerable loss of quality. Various compression techniques have been used by researchers to compress speech signal. This paper deals with the study of different speech compression techniques based on discrete orthogonal transforms. Here we present the different approaches adopted by different authors to compress speech signal based on orthogonal transform such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Discrete Walsh Hadamard Transform (DWHT) and Wavelet Packet Decomposition (WPD).

To Cite This Article: N. Vini Antony Grace, Dr. A. Chilambuchelvan, Dr. N. Usha Bhanu, A Comprehensive Survey on Speech Compression Techniques Based on Discrete Orthogonal Transforms. ***Aust. J. Basic & Appl. Sci.,*** *9(16): 229-234, 2015*

1 INTRODUCTION

Speech compression is a ripe subject of research. It is the process of obtaining a compact representation of voice signals for efficient transmission over band limited wired and wireless channels and / or storage. A speech compression system focuses on reducing the amount of redundant data while preserving the integrity of signals.

Compression techniques can be broadly classified into two classes: Lossless and Lossy compression. In lossless compression the original signal can be perfectly recovered from the compressed signal. In case of lossy compression, the original signal cannot be perfectly recovered from the compressed signal, but gives its best possible quality for the given technique. Lossy compression typically attain far better compression ratio than lossless compression by discarding less critical data.

The compression methods can be classified into three functional categories: the first is compression by direct method; the samples of the speech signal are directly handled to provide compression. The second is compression by transformations such as Discrete cosine Transform, Discrete Fourier Transform, Discrete Wavelet Transform etc., The third method is compression by parameters extraction; the input signal is analyzed to extract some parameters that are used later to reconstruct the signal.

During the last decade, wavelet transform, more particularly DWT has emerged as a powerful and robust tool for analyzing and extracting information from non-stationary signal due to time varying nature of these signals. The primary objective of this paper is to summarize the different methods that have been developed based on wavelet or wavelet packets for speech compression. Also a comparative study of the performance of the wavelet transform and other orthogonal transforms are presented.

This paper is organized as follows. Section 2 presents the general procedure used for transform based speech compression. Section 3 discusses the literature survey of various speech compression algorithms using different transforms, its pros and cons. In addition a comparative study between the performances of the transforms is also presented. Section 4 talks about some of the objective performance measurement parameters used for speech compression. In section 5, the review results are discussed and finally conclusions are drawn in section 6.

Corresponding Author: N. Vini Antony Grace, Part – Time Research Scholar-Anna University, Chennai

Australian Journal of Basic and Applied Sciences, 9(16) Special 2015, Pages: 229-234

2 Transform Based Speech Compression General Procedure:

The idea of speech compression using transform method is primarily related to the representation of a speech signal in the form of transformed coefficients. As a result of applying transform on the signal many of the coefficients will either be zero or have negligible magnitudes. Most of the signal energy will be concentrated in the high valued coefficients, which are very few. Consequently the speech compression is achieved by truncating the small valued coefficients and efficiently encoding them. A typical block diagram of speech compression based on transforms is shown in Fig 1.Following are the general procedure for speech compression using transforms.

2.1 Speech coding algorithm:

Step 1: Choosing the frame size:

Choose an appropriate frame size and divide the original speech signal into frames of chosen size. This is important because framing aim to improve the compression ratio obtained.

Step 2: Applying the transform technique:

Apply the transform [DCT/ DWT/ DWHT/ WPD] on each frame of the speech signal to extract the coefficients. In case of DWT we have to choose an appropriate decomposition level and mother wavelet before applying the transform. The choice of the mother wavelet is important because it directly affects the SNR of output signal by maximizing it and minimizing the relative error.

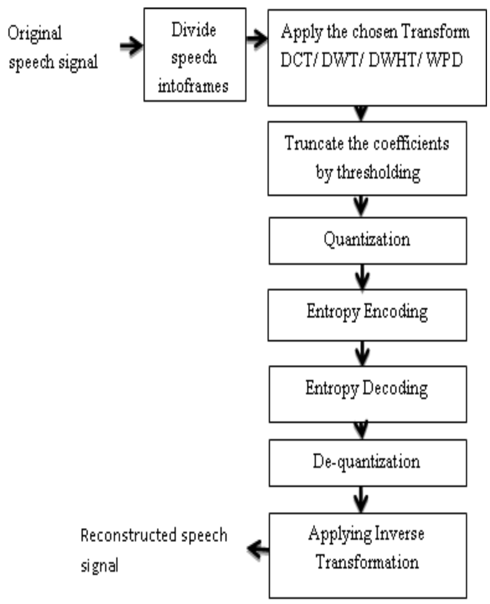

Fig. 1: General Procedure for speech compression.

Step 3: Thresholding:

The coefficients obtained after applying the transform on the frame concentrate energy in few neighbors. Thus we can truncate all coefficients with low energy value and retain few coefficients with high energy. Two different thresholding techniques that are used in common are

1. Global thresholding

This technique works by retaining the transform coefficients which have the largest absolute value.

2. Level Dependent thresholding

This technique is derived from the Birge-Massart strategy. The value of the threshold applied depends on the compression ratio we want to achieve.

Step 4: Quantization:

It is a process of mapping a set of continuous value input data to a set of discrete valued output data. Aim of quantization step is to decrease the information found in threshold coefficients in such a way that the quantization process produces no error. There are many quantization methods such as: uniform, scalar and vector quantization.

Step 5: Encoding:

The quantized coefficients for each sub band contain some redundancy which is a waste of space. To remove it entropy coding like Huffman coding or arithmetic coding is used.

2.2 Speech decoding algorithm:

For reconstructing the speech signal the reverse process of the speech coding has to be done. It can be summarized in three steps.

Step 1: Entropy decoding of the quantized sub bands.
Step 2: De quantization of the encoded sub bands.
Step 3: Applying the inverse transforms

3 Literature Survey:

To study and analyze more about speech compression techniques, the following literature survey has been done and discussed in this chapter. Because of the limitations on the bandwidth and storage capacity, speech signal needs to be compressed before the transmission process or the storage process. There are many speech compression techniques evolving every day. Hence it is necessary to study a literature about it, to understand the techniques and to use the appropriate methods during compression of speech signals.

(Kumar, 2013) In this paper, the authors have proposed optimized wavelet filters for speech compression. The filter coefficients are optimized by deriving it with Blackman and Kaiser Windows via linear optimization. Developed wavelets are used as mother wavelets and 5 level decomposition of DWT is applied on speech signal. Global thresholding is used and the coefficients above threshold are quantized using uniform step size. Huffman encoding is used to encode the truncated small valued coefficients. A comparative study of performance of different existing wavelet filters like Haar wavelet, Db10 wavelet and the proposed wavelet filter is made in terms of compression Ratio (CR), Signal to Noise Ratio(SNR), Peak Signal to Noise Ratio(PSNR) and Normalized Root Mean Square Error(NRMSE). It has been shown the quality of reconstructed signal of the proposed wavelet filters is far superior to other existing wavelet filters. The developed wavelet filters also yields comparable compression ratio.

(Preet Kaur 2012) In this paper, authors have exploited the techniques of Wavelet Packet Decomposition (WPD) and Discrete Wavelet Transform (DWT) to materialize speech compression. The performances of these two techniques are compared in terms of SNR, PSNR, NRMSE and retained Signal Energy (RSE). In both DWT and WPD three different wavelet filters Haar, db2 and db4 are used and 6 level decomposition is applied. Both global and level dependent thresholding are implemented separately for all the three wavelets and Huffman encoding is used to compress the speech signal. From the results it has been proved that SNR of all the three wavelets are good when WPD is used compared to DWT. This method also gives a comparable compression ratio as that of ordinary DWT. The performance parameters were found to be better in global thresholding.

(Siva Nagu, 2012) In this paper authors have jointly proposed Adaptive Kalman Filtering method along with wavelet coding to filter signals from noise. The purpose of this approach is to reconstruct an output speech signal with perfect audibility by making use of accurate estimating ability of the Kalman filter. English words have been used for his experiment and the code was simulated using MATLAB. The proposed wavelet coder was tested on male and female speech signals of duration 10sec with wavelets like HAAR, sym2, sym5, coif2 & db20. The results obtained were compared with the results of ordinary wavelet coding technique. It was observed that for all the wavelets mentioned above the Kalman filter method with wavelet coding provides better Peak signal to Noise ratio(PSNR) than wavelet coding.

(Noureddine Aloui, 2013) A new lossy speech compression algorithm based on Discrete Walsh HadamardTransform (DWHT) for stationary signals was proposed in this paper. Here the authors have compared the performance of DWT & DWHT in terms of CR, SNR, PSNR, NRMSE & CPU time. Speech signal is divided into frames of size 256 and global thresholding is applied for DWHT. For DWT algorithm the used mother wavelet is db10. The signal is decomposed till level 5 and global thresholding is applied. A comparative study between the proposed algorithm and the DWT algorithm prove that the proposed algorithm outperforms the algorithm based on DWT in terms of: CPU time, CR, SNR and NRMSE. The total CPU time used by MATLAB for running DWHT algorithm is decreased to about 65.5% compared to DWT algorithm. However the quality of the reconstructed signal using the proposed algorithm is degraded for large frame size. So the authors conclude that this new compression algorithm based on DWHT can be efficiently used for speech compression with stationary frame.

(Vithalani, 2003) In this paper, instead of classical sub band coding method, the authors have implemented lifting scheme of wavelet transform for the purpose of speech compression. This is because compared to traditional sub band coding method this lifting scheme requires fewer computation and easy to implement for real time low power applications. Here Daubechies D4 wavelet is factored into lifting steps and the signal is decomposed into coarse and detail parts. The detail coefficients below certain threshold are reduced to zero and then standard entropy coding is used for speech compression. The signal is reconstructed through decoding and an inverse transform. The algorithm is implemented using programming language C and visual C++. To simulate real time signals, entire file is divided into blocks of 256 bytes. Each block is processed separately. Total computation time for each block is found to be less than 1ms with CPU speed of 600MHz. Compression Ratio (CR) with different threshold were taken. From the results of Mean Opinion Score it is observed that for threshold equal to zero it provides loss less compression but as threshold increases the quality of speech signal degrades.

(Harmanpreet Kaur) An efficient hybrid combination of DWT-DCT techniques for compression of speech has been proposed in this paper. The original signal is first compressed by DWT and compressed signal is again compressed by DCT. The compressed signal is decompressed by DWT. The performance of the signal is measured by PSNR and MSE using different wavelet filters like Daubechies, Symlets, Biorthgonal and Coiflet. The result shows that the PSNR and MSE of signal are improved much in this hybrid model. With this high efficiency of compression can be achieved.

(Smita Vatsa, 2012) In this paper, DWT and DCT transform based speech compression techniques are implemented with Run Length Encoding (RLE), Huffman encoding and RLE followed by Huffman encoding. A comparative analysis of performance of transform technique based speech compression systems for 3 different encoding methods is done. Global thresholding is applied for DCT coefficients and level dependent thresholding based on Birge-Massart strategy is applied for DWT (Db10) coefficients. After thresholding the coefficients are quantized using uniform quantization and then the process of encoding is done using 3 different approaches. From the experiments it has been concluded that when in transform based speech compression system if DWT with run length followed by Huffman encoding is exploited the speech signal can be represented with minimum data values and the reconstructed speech obtained is also intelligible.

(Noureddine Aloui, 2015) An optimized speech compression algorithm using DWT and its real time implementation on fixed point digital signal processor is presented in this paper. Optimization is done by adding a Voice Activity Detector (VAD) module before the application of DWT. The VAD module avoids the computation of DWT coefficients during the inactive voice signal. The speech signals are extracted from the TIMIT database and analyzed by frames of 256 samples. For each frame, the VAD module provides decision, which is used to switch between the speech coding using DWT or silence coding. For DWT coding the used mother wavelet is Db10 and the level of decomposition is 5. Performance of this method of speech coding with VAD module is experimentally proven to have much improvement than speech coding without VAD module in terms of CPU time, CR, SNR, PSNR and NRMSE. This optimized speech compression algorithm using DWT is then implemented using the fixed point processor TMS320C6416 DSP for processing real time signals. A comparative study of performance between the optimized version and original version of DWT speech coding algorithm using the processor is made with some objective criteria such as the cycles count required to perform each algorithm and the memory sizes of DSP executable code. From the results it can be observed that the total number of cycles decreases to about 49.6% by using the VAD module. Thus the authors have proved that the addition of the voice activity module to the speech codec based on wavelets reduce s the complexity of computation, the bit rate and achieves high speech coding performances.

(Noureddine Aloui, 2015) A new algorithm for designing Quadrature Mirror Filters (QMF) banks using windowing technique has been proposed in this paper. The designed QMF banks are used as mother wavelet for speech compression algorithm based on DWT. In the proposed algorithm the window type(Blackman, Hanning, Barlett, Kaiser…), the filter order (N), the pass band ripple(A_p), the stop band attenuation(A_s), the tolerance(tol), three cut off frequencies(ωc_1, ωc_2, ωc_3) and the magnitude response in the ideal condition(MRI=0.707) are all initialized. This algorithm is developed in MATLAB language and used for iteratively optimizing the cut off frequency. The iterations are stopped only when the tolerance is satisfied (m <tol), in this case the cut off frequency (ωc) is defined as the minimum of the cut off frequencies $\omega c_{i/i=1,2,3}$. Different mother wavelet filters (optimized Hanning, Kaiser & Triangular…) are designed using the proposed algorithm based on windowing techniques and 5 level decomposition of DWT are applied after dividing the speech signal by frames of 256 samples. The obtained sub bands are thresholded using global threshold and hard threshold. The obtained wavelet coefficients after thresholding are quantized, multiplexed and coded. The reverse process is done for reconstruction. Three tests were performed to evaluate the efficiency of the developed algorithm. The first test is an evaluation of optimized wavelet filters in the field of speech compression. The second is a comparative performance study between the developed wavelet filters and Daubechies mother wavelets. The third is a comparative study between the proposed algorithm with other existing algorithms given in [Creusere ***et al,*** Jin ***et al &*** Kumar ***et al*** (2010)], used to design QMF banks for some specifications; Reconstruction Error (RE) and No. of iterations (NI). The tabulated results in the paper clearly shows that the optimized wavelet filters are able to compress speech signal, when they are used as mother wavelets for speech compression algorithm based on DWT. The performance parameters CR, SNR, PSNR and NRMSE depend on the used window type. From the comparative performance study between the optimized QMF banks and the classical Db wavelets, it can be observed that the global performances are significantly improved by using the optimized wavelet filters. The comparison results with other existing algorithms used for designing QMF bank show an important reduction in terms of RE and NI.

4 Performance Measurement Tools:

A number of quantitative parameters can be used to evaluate the performance of the coder in terms of both the reconstructed signal quality and compression scores. Most commonly used parameters are:

Signal to Noise Ratio (SNR):

$$SNR = 10\, log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) \qquad (1)$$

$\sigma_x^2$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):

$$PSNR = 10log_{10}\frac{NX^2}{\|x-r\|^2} \qquad (2)$$

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 difference between the original and reconstructed signal.

Normalized Root Mean Square Error (NRMSE):

$$NRMSE = \sqrt{\frac{\left(x(n)-r(n)\right)^2}{\left(x(n)-\mu(n)\right)^2}} \qquad (3)$$

x (n) is the speech signal, r (n) is the reconstructed signal and μ (n) is the mean of the speech signal.

Australian Journal of Basic and Applied Sciences, 9(16) Special 2015, Pages: 229-234

Retained Signal Energy (RSE)

$$RSE = \frac{100 * \|x(n)\|^2}{\|r(n)\|^2} \quad (4)$$

$\|x(n)\|^2$ is the norm of the original signal and $\|r(n)\|^2$ is the norm of the reconstructed signal.

Compression Ratio (CR)

$$CR = \frac{Number\ \ of\ significant\ \ transform\ \ coefficients}{Total\ number\ \ of\ transform\ \ coefficients} \quad (5)$$

5 Review Results and Discussions:

We have reviewed the various speech coding techniques based on different orthogonal transforms such as Discrete Cosine Transform, Discrete Wavelet Transform, Discrete Walsh Hadamard Transform, Wavelet Packet Decomposition and their experimental outcomes given by the respective authors. Experimental comparisons on different transforms suggest a recipe described as follows: 1) Wavelet filters derived with Blackman and Kaiser Windows performs exceptionally well compared to other existing wavelet filters. 2) If the performance of DWT and WPD are compared the SNR and CR are found to be good in WPD. 3) Adaptive Kalman filtering if used with wavelet coding provide better filtering of noise from signals. Thereby it improves the PSNR of signal than ordinary wavelet coding. 4) The performance parameters are found to be good in global thresholding than level dependent thresholding in all the transform techniques. 5) The DWHT algorithm outperforms the DWT algorithm for speech signals with stationary frame. However for real time signals with large frame size it is not satisfactory. 6) Lifting scheme of wavelet transform also yields comparable results. 7) Reconstructed speech obtained is intelligible if DWT is exploited with RLE followed by Huffman encoding. 8) Addition of Voice Activity module before the application of DWT reduces the complexity of computation, the bit rate and achieves high speech coding efficiency. 9) If QMF banks designed using windowing techniques are used as mother wavelets in speech compression algorithm based on DWT, the efficiency increases in terms of reduction in RE and NI.

6 Conclusions:

Quality of speech after compression is the main criteria that all compression techniques should hold. An extensive literature survey on speech compression algorithms based on Discrete Orthogonal Transforms is performed in this paper. We have mainly focused on four different transforms Viz. DCT, DWT, DWHT and WPD. Many algorithms are proposed by different authors by performing some variations on the basic idea of the transform technique to provide better result. Though there are many algorithms proposed with unique characteristics, research has to be done to develop techniques that will produce high quality reconstructed speech signal with high compression ratio and enable their use in portable and mobile devices which have limited computing power.

REFERENCES

Creusere, C. and S. Mitra, 1995. "A Simple Method for Designing High quality Prototype Filters for M- Band Pseudo QMF Bank," *IEEE Transactions on signal processing*, 43(5): 1005-1007.

Harmanpreet Kaur, Ramanpreet Kaur, "Speech Compression and Decompression using DWT and DCT". *International Journal of Computer Technology & Application*, 3(4): 1501-1503.

Jin, A., R. Saxena and C. Saxena, 2005. "A Simple Alias Free QMF System with Near Perfect Reconstruction," *Journal of Indian Institute Science*, 85(1): 1-10.

Kumar, A., G. singh and R. Anand, 2010. "Design of Quadrature Mirror Filter Bank using Particle Swarm Optimization," *International Journal on Electrical and Power Engineering*, 1(1): 41-45.

Kumar, A., G.K. Singh, G. Rajesh, K. Ranjeet, 2013. The optimized wavelet filters for speech compression. *International Journal of speech Technology (Springer)*, DOI 10.1007/s10772-012-9173-1.

Noureddine Aloui, Chafik Barnoussi, Adanecherif, 2015. "A New algorithm for QMF Banks Design and Its Application in Speech Compression using DWT". *International Arab Journal of Information Technology*, 12(1).

Noureddine Aloui, Souha Bousselmi, Adanecherif, 2013. "Speech Compression Based on Discrete Walsh Hadamard Transform". *International Journal of Information Engineering and Electronic Business*, 3: 59-65.

Noureddine Aloui, Souha Bousselmi, Adanecherif, 2015. "Optimized Speech Compression Algorithm Based on Wavelets Techniques and its Real Time Implementation on DSP". *International Journal of Information Engineering and Electronic Business*, 03: 33-41.

Preet Kaur and Pallavi Bahl, 2012. "Comparative Analysis between DWT and WPD Techniques of Speech Compression". *IOSR Journal of Engineering (IOSRJEN)* ISSN. 2250-3021, 2(8): 120-128.

Siva Nagu, T., K. Jyothi, V. Sailaja Speech, 2012. "Compression for Better Audibility using Wavelet Transformation with Adaptive Kalman Filtering". *International Journal of Computer Applications*,53(11).

Smita Vatsa, O.P. Sahu, 2012. "Speech Compression Using Discrete Wavelet Transform and Discrete Cosine Transform". *International Journal of Engineering Research &Technology (IJERT)*, 1(5): 2278-0181.

Vithalani, C.H., H.B. Dave, 2003. "Wavelet Based Speech Codec". *International Symposium on*

Intelligent Signal Processing and Communication Systems(ISPACS 2003), pp: 7-10.