NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 2, ISSUE 12, DEC.-2015

SPEECH COMPRESSION TECHNIQUES: A REVIEW

Pujari Bhavana C.
ME Student
Amrutvahini College of engineering,Sangamner

ABSTRACT

Speech is the vocalizer form of human communication,and based upon the syntactic combination of lexical and vocabularies. The aim of speech coding is to compress the speech signal to the highest possible compression ratio but maintaining user acceptability.There are many methods for speech compression like Linear Predictive coding (LPC) , Code Excited Linear Predictive coding (CELP) ,Sub-band coding ,Transform coding :- Fast Fourier Transform (FFT) ,Discrete Cosine Transform (DCT) , Continuous Wavelet Transform (CWT) , Discrete Wavelet Transform (DWT) ,Variance Fractal Compression (VFC) , Discrete Cosine Transform (DCT),Psychoacoustics andetc. Few of them are discus in this paper.

KEYWORDS: Compression, LPC,DWT,DCT.

INTRODUCTION

Speech compression is nothing but reduction of number of bits needed to represent the signal used for storage purpose and transmission. The ideal goal of speech compression is to contain original information in as minimum bits as possible.The reasons for compressing the signal is Cost of disk, Cost of data management,Memory,Bandwidth and transfer speed.There are two basic types of compression lossy and lossless.

LOSSLESS COMPRESSION:-

In this type of compression signal after compression is same as before,no information has been loosed i.e. the original signal can be perfectly recovered from the compressed signal. It is mainly used in application where it is necessary that the original signal and the de-compressed signal are almost same.

Examples: Entropy Encoding (Shannon-Fano Algorithm, Huffmann coding, Arithmetic Coding) Run-length, Lempel Ziv Welch (LZW) Algorithm.

Lossy compression:-In this type of compression,some degree of information has been lossed. The original signal cannot be perfectly recovered from the compressed signal,but it gives its best possible quality for the given technique. Lossy compression typically attain far better compression than lossless by discarding less-critical data. Theaim of this technique is to minimize the amount of data that has to be transmitted.They are mostly used for multimedia data compression.

Ex: FFT,DCT,DWT.

1.1 Linear Predictive coding (LPC):

LPC is most commonly used in speech coding due to effectiveness of LPC coefficient in modelling vocal tract associated with speech production.LPC is used to estimate basic speech

parameters like pitch formant and spectra.The principle behind the use of LPC is to minimize LPC coefficient .This LPC coefficient is estimated in energy frame size of 20ms long.
LPC analysis of each frame involves decision making process of concluding if sound is voiced or unvoiced.If sound is decided to be voiced,on impulse train is used to represent it with non zero taps occurring every pitch period .Autocorrelation function is one of technique used to estimate pitch period.For unvoiced frame white noise is used to represent it and pitch period of T=0 is transmitted.

1.2 Discrete Cosine Transform (DCT):

DCT forming a periodic,symmetric sequences from finite length sequence in such a way that original finite length sequence can be uniquely recovered.It can be used for speech compression because of high similarities in adjacent coefficient.DCT is similar to DFT but containing only the real part of DFT.
In speech processing DCT

The 1D DCT is

$$Y(k)=w(k)\sum_{n=1}^{N} x(n) * \cos\left(\frac{\pi*(2n-1)*(k-1)}{2N}\right)$$

K=1,2,3......N

Where $w(k)=\frac{1}{\sqrt{N}}$ k=1

$=\sqrt{\frac{2}{N}}$ 2<=k<=N

N is the length of x
X and y are of same size
For reconstruction very few DCT coefficient are required.

$$x(n)=\sqrt{\frac{2}{N}}*\sum_{k=0}^{N-1} w(k) * x(k) * cos\frac{(2n+1)n\pi}{2N}$$

1.3 Discrete Wavelet Transform (DWT):

DWT is special property of wavelet transform that provide a compact representation of signal in time and frequency domain.DWT decomposes the signal into the too many function by using property of translation and dilation of single function called as a mother wavelet.

$$\varphi_{s,\tau} = \frac{1}{\sqrt{s}} * \varphi\left(\frac{t-\tau}{s}\right)$$

Where s is scaling parameter
$\tau$is translation parameter
DWT of signal s(k) is defined as

$$DWT(m,n)=2^{-m/2}*\sum_k s(k) * \varphi(2^{-m}\, k-n)$$

DWT is sub band coding based technique.
In DWT signal which is to be analysed is first passing through filter bank followed by decimation operation .This filter bank consist of LPF and HPF at each decomposition stages.
LPF O/P is called approximate component
HPF O/P is called detail component
Working of DWT is as shown in figure 1.1

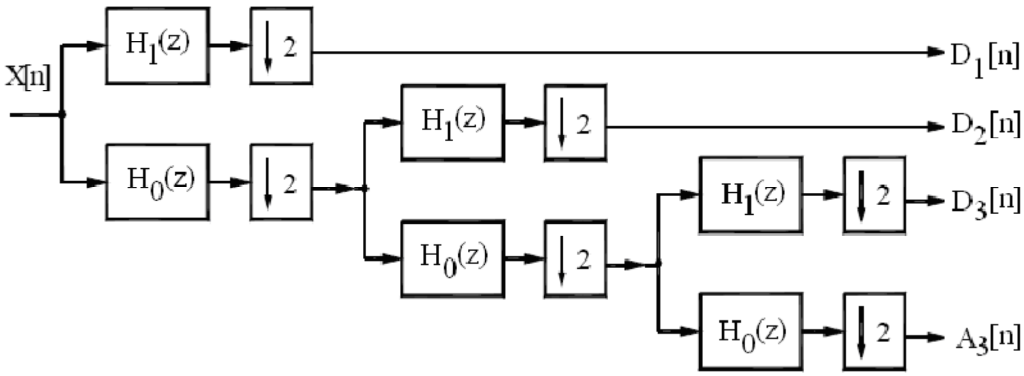

Figure 1.1 Three-level wavelet decomposition trees

1.4 Discrete Wavelet Packet Transform:

In this signal is split into approximate and detail coefficient then both the coefficient is then itself split into second level approximate and detail coefficient and process is repeated ,as shown in Figure 1.2

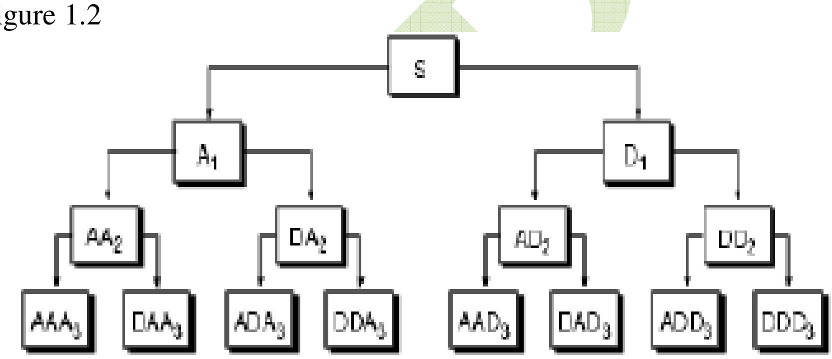

Figure 1.2 Level 3 Decomposition using Wavelet Packet Transform

It gives more than 2^{2n-1} different ways to encode the signal.

The wavelet have several families, they are Haar, Daubechies, Symlet, Coiflet, Biorthogonal, Reverse Biorthogonal, Meyer wavelet, Gaussian, complex Gaussian, Maxican Hat, Morlet, Complex Morlet, Ballet Lamarie.

1.5 Psychoacoustic Model:

It is based on study of human perception.The average human hearing of all frequency is not same. Psychoacoustic Model is made up of two principal human auditory system properties,they are auditory masking and hearing absolute threshold.It uses the concept that some informationin signal is not necessary for our interpretation of sound,thus they can be removed.The speech signal contains lots of frequency many of whom the human ear can't hear .By removing these frequency from the signal,the information load gets reduced without effecting our impression of signal.

1.5.1 Frequency Masking:

It occurs when frequency we able to hear normally is masked by nearby frequency. The ear is unable to simply distinguish frequency close to each other .The masked frequency can be removed.

1.5.2 Temporal Masking:

When weak frequency is preceded by a strong frequency in time domain,that is frequency with low energy close to a frequency with high energy,the sound associated with weak frequency is unable to hear if time interval between frequencies is short.This is called temporal masking.By removing all frequency that are masked,the ones with low energy the information amount is minimized.

CONCLUSION

From review of speech compression techniques, it is observed that, the greatest advantage of wavelet over other techniques is that the compression factor is not constant and it can be varied while most other techniques have fixed compression factor. DWT significantly improves the reconstruction of the compressed speech signal and also yields higher compression factor.

REFERENCES

[1]. Jing Pang, Shitalben Chauhan, Jay Mahesh kumar Bhlodia," *Speech Compression FPGA Design By Using Different Discrete Wavelet TransformSchemes", in Advances in Electrical and Electronics Engineering - IAENG Special Edition of the World Congress on Engineering and Computer Science 2008.*

[2]. Shijo M Joseph , Babu Anto P," *SPEECH COMPRESSION USING WAVELET TRANSFORM", in IEEE-International Conference on Recent Trends in Information Technology, ICRTIT 2011 ,MIT, Anna University, Chennai. June 3-5, 2011.*

[3]. Firoz Shah A, Babu Anto P," *Spoken Digit Compression: A Comparative Study between Discrete Wavelet Transforms and Linear Predictive Coding", in International Journal of Computer Applications (0975 – 8887) Volume 6– No.6, September 2010.*

[4]. Jithin James,Vinod J Thomas,"*A Comparative Study of Speech Compression using Different Transform Techniques", in International Journal of Computer Applications (0975 – 8887) Volume 97– No.2, July 2014.*

[5]. Harmanpreet kaur , Ramanpreet kaur," *Speech compression and decompression using DWT and DCT" in Harmanpreet Kaur et al ,Int.J.Computer Technology &Applications,Vol 3 (4), 1501-1503.*