SPEECH COMPRESSION ANALYSIS USING MATLAB

Manas Arora[1], Neha Maurya[2], Poonam Pathak[3], Vartika Singh[4]

[1, 2]M.Tech Student, Digital Communication, BBDNIIT, Uttar Pradesh, India,
[3]Associate Professor, [4]Sr. Lecturer, Electronics & Comm., BBDNIIT, Uttar Pradesh, India

Abstract

The growth of the cellular technology and wireless networks all over the world has increased the demand for digital information by manifold. This massive demand poses difficulties for handling huge amounts of data that need to be stored and transferred. To overcome this problem we can compress the information by removing the redundancies present in it. Redundancies are the major source of generating errors and noisy signals. Coding in MATLAB helps in analyzing compression of speech signals with varying bit rate and remove errors and noisy signals from the speech signals. Speech signal's bit rate can also be reduced to remove error and noisy signals which is suitable for remote broadcast lines, studio links, satellite transmission of high quality audio and voice over internet This paper focuses on speech compression process and its analysis through MATLAB by which processed speech signal can be heard with clarity and in noiseless mode at the receiver end .

Keywords: *Speech compression, bit rate, filter, MPEG, DCT.*

---***---

1. INTRODUCTION

Data compression is a technique in which data content of the input signal to system is compressed so that original signal is obtained as output and unwanted or undesired signals are removed. Therefore when speech signals are used in the form of data it is termed as SPEECH COMPRESSION.

Speech is a very basic way for humans to convey information to one another. Speech has a small bandwidth of 4 kHz. Speech compression involves coding of real-time audio signals at the lowest possible bit rates. The compression of speech signals has many practical applications. It is used in digital cellular technology where many users share the same frequency bandwidth. Compression allows more users to share the system at a particular time. It can also be used for digital voice storage that are used for answering machines and pre-recorded telephone calls that are used for purpose of providing any kind of information to user or advertising. For a given memory size, compression allows longer messages to be stored than otherwise [1].

2. SPEECH COMPRESSION

Speech compression enables efficient storage and transmission of data. There may be varying amounts of compression in data according to the sampling rate used. This gives different levels of system complexity and compressed quality of speech data. The recorded waveform which is compressed can be transmitted with or without loss. Therefore, there are two types of compression: lossy and lossless. The digital audio data is processed through mixing, filtering and equalization. The speech signal is fed into an encoder that uses fewer bits than original audio data bit rate. This results in reducing the transmission bandwidth of digital audio streams and also reduces storage size of audio files. [15] Compression may be lossy or lossless. Lossy compression is transparent to human audibility but lossless being have a compressing factor from 6 to 1. [15].An uncoded speech signal is [17]

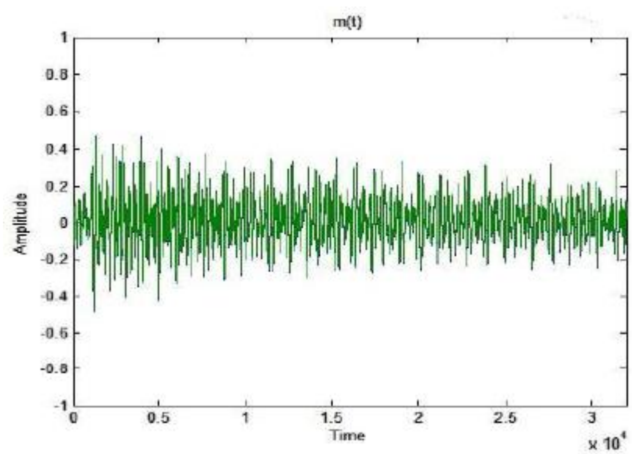


Fig 1: Uncoded speech signal

At the decoder, bit stream decoding bit is done which is followed by frequency sample reconstruction. Finally frequency-to-time mapping produces decoded speech signal. DCT is used to compress speech signals. [15].

Discrete Cosine Transform (DCT) is a main transform or mapping method used which maps the time domain into frequency domain. A DCT expresses a finite sequence of data points in terms of a sum of cosine functions oscillating at different frequencies. The discrete cosine transform is a linear, invertible function defined as F: RN →RN (where R denotes the set of real numbers), or equivalently an invertible N × N square matrix. Mathematically DCT is calculated from (1) [2]

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\Pi\left(n+\frac{1}{2}\right)k}{N}\right]$$

$$k = 0,......N-1 \qquad (1)$$

The cosine graphs define the nature of frequency used in DCT transform and define the frequency of the speech signal. The cosine graphs define the window function. The window function has cosine function as (2).

$$\cos\left[\frac{\Pi\left(n+\frac{1}{2}\right)}{N}\right] \qquad (2)$$

Basic cosine graphs used in the discrete cosine transform are:

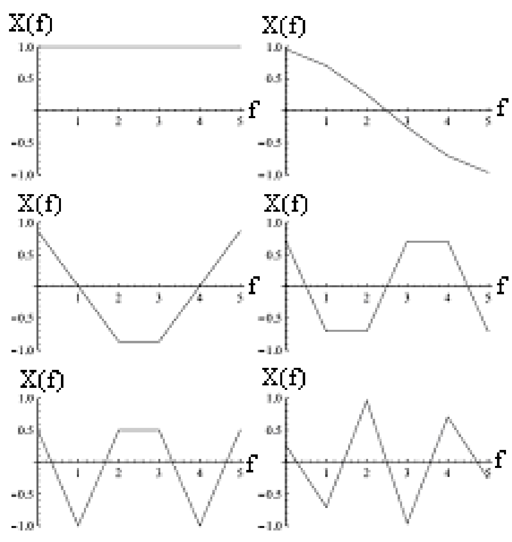


Fig 2: Cosine graphs used in DCT

The speech compression using DCT method is an efficient method by which we can modify the frequency of the signal and analyze the signals in different frequency ranges in a fixed frequency domain. DCT also effectively compresses the speech signals and maintain the audibility feature of speech signals.

3. NEED OF SPEECH COMPRESSION

A basic question arises what is the need of speech compression? Answer lies within the requirement of appropriate bit rate for high quality speech signal. Bit rate is defined as number of sample per second which is given by Rbits calculates from (3) where fs is the sampling frequency and n describes the number of bits.

$$R_{bits} = f_S \times n \qquad (3)$$

The requirement of achieving high bit rate and clarity of audio signals basically defines the need of speech compression[14]. A given frequency bandwidth can be used by a number of users at a time and moreover, larger size speech data files can be stored.

4. SPEECH COMPRESSION TECHNIQUE IN MPEG TECHNOLOGY

A speech signal is a complex signal which has a wide spectrum and MPEG helps in perceptual phenomena for our ears. The large frequencies and smaller frequencies below masking threshold are inaudible to human ears. MPEG standard helps in defining standard and filtering audio signals in the available bandwidth and maintains signal to quantization ratio. In MPEG-1 audio compression is performed as given in figure 1 and involves two process. Filter bank which uses filters divides spectrum of incoming signals in sub bands. The quantizer quantizes the sub bands. [10]

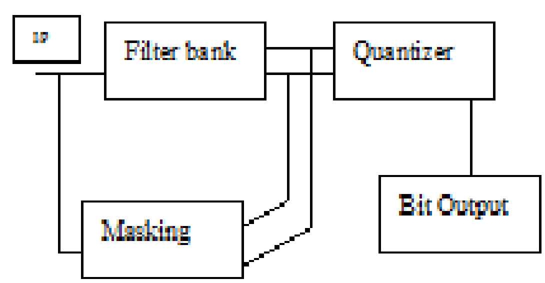


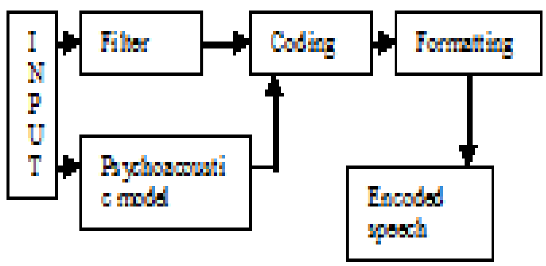
Fig 3: Block Diagram of MPEG-1 Encoder

The basic method by which speech compression in MPEG technology is done is by utilizing an MPEG encoder and decompression is performed by MPEG decoder. Figure 3 shows block diagrams of the MPEG/speech encoder and decoder. [11, 12]

In MPEG, compression of the input stream passes through a filter bank which divides the input into multiple sub bands. The input is simultaneously also passed through a psychoacoustic model which is used to determine the signal-to-mask ratio of each sub band. The bit or noise allocation block utilizes the output of psychoacoustic model and a decision is made about the total number of code bits which are available for quantization process [3, 13].

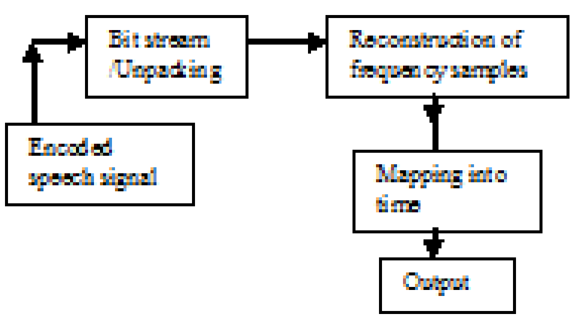
Code bits also help in removing quantization noise. The output of compressor is quantized audio samples and these formats of the data are transformed into a decodable bit stream [3, 13].

In MPEG, decompression of the speech signals simply reverses the formatting and then reconstruction of quantized code bits is done then reconstructs the quantized sub band values, which are finally transformed into the set of sub band values into a time-domain audio signal [3, 13].



SPEECH ENCODER

Fig 4: Speech Encoder block diagram



SPEECH DECODER

Fig 5: Speech Decoder block diagram

The MPEG speech standard has three distinct layers for compression.

- Layer I is the basic algorithm
- Layers II and III are enhancements of Layer I. Each successive layer improves the compression performance at the increasing cost of encoder and decoder complexity [14].

4.1 Layer I:

The Layer I algorithm uses the basic filter bank which is found in all layers. This filter bank divides the audio signal into 32 constant-width frequency bands. The filters are simple and provide time and frequency resolutions by which it is easily perceived by human ears. The design of Layer I has three concessions. First, the 32 constant width bands which do not accurately reflect the ear's critical frequency bands. Second, the filter bank and its inverse are not lossless transformations. Thirdly, adjacent filter bands have a significant frequency overlap [4, 5, 6, 7, 8, 15]

4.2 Layer II:

The Layer II algorithm is a simple enhancement of Layer I. It improves compression performance by coding data in larger groups. By the Layer II encoder forms frames of 3 by 12 by 32 that is total of 1152 samples per audio channel as compared to Layer I codes data in single groups of 12 samples for each sub band .Layer II removes stereo redundancy coding and there is one bit allocation The encoder encodes with a unique scale factor for each group of 12 samples to avoid audible distortion. The Layer II algorithm improves performance as it uses efficient code for representing the bit allocation, the scale factor values, and the quantized samples. [4, 5, 6, 7, 8, 15]

4.3 Layer III:

The Layer III algorithm is a much more refined approach.[16,17] though it is based on the same filter bank found in Layers I and II. Layer III compensates for some filter bank deficiencies by processing the filter outputs with a modified discrete cosine transform (MDCT) [4, 5, 6, 7, 8, 15].

These all the three layers completely design the methods of audio compression in MPEG technology.

5. IMPLEMENTATION IN MATLAB

MATLAB is a useful tool which is used to analyze speech signals which are read in >wav format. Following commands are used to analyze.

- ***wavread:*** it reads speech signal
- ***windowsize:*** defines window function of transformation
- ***wavplay:*** it produces speech signal after transformation
- ***length:*** defines length of speech to be processed by transforming principle
- ***dct:*** performs discrete cosine transform
- ***idct:*** performs inverse discrete cosine transform

6. OUTPUT IN MATLAB

In speech signal analyzing through MATLAB we obtain following spectrums

- Speech Signals with different Amplitudes(as input data)
- Portion of signal according to length and window-size
- Speech spectrograms

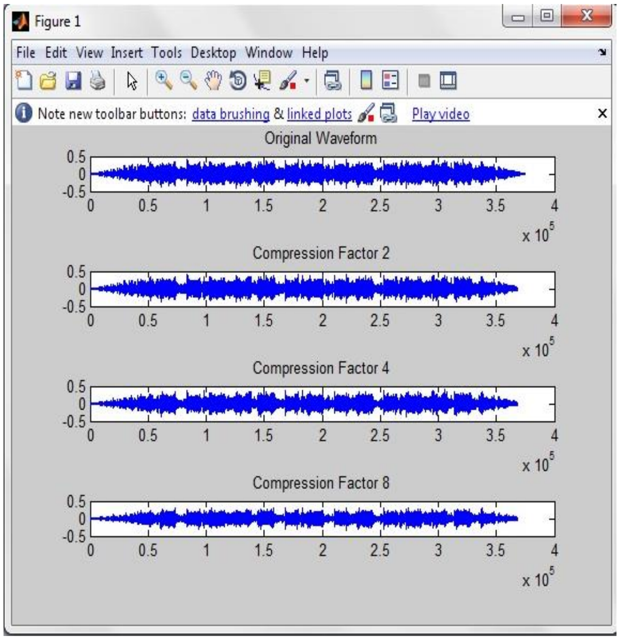


Fig 6: Speech Signals with different Amplitudes

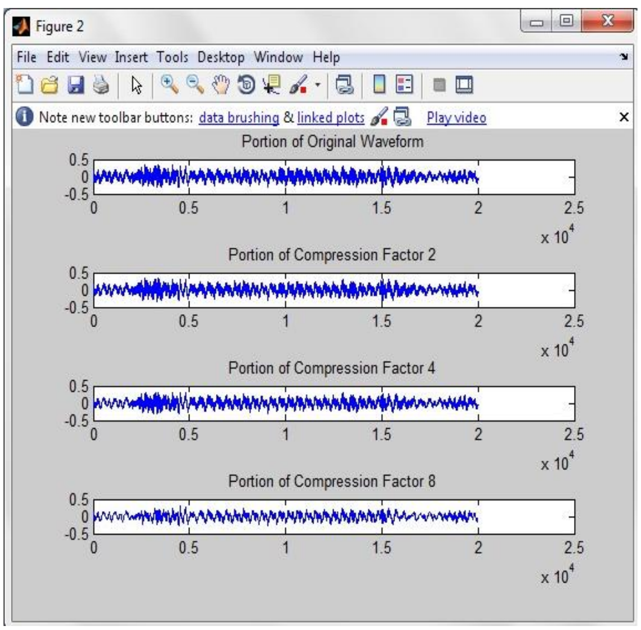


Fig 7: Portion of speech signal to be processed

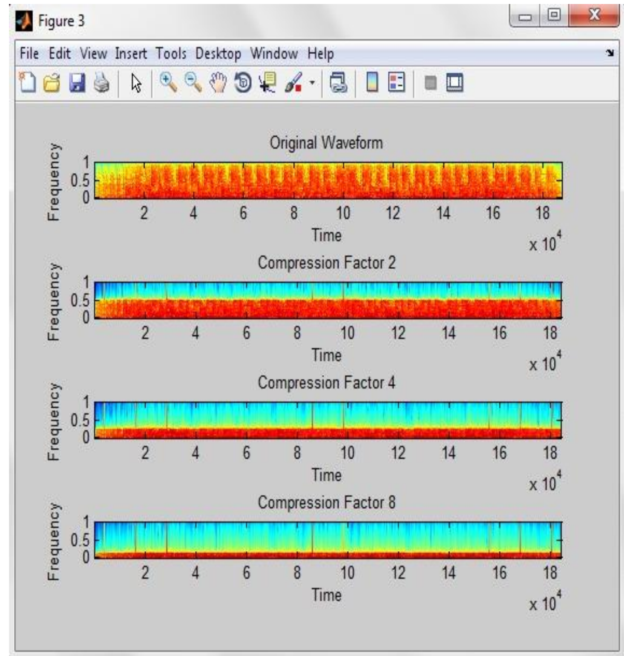


Fig8: Speech Spectrograms

7. PRESENT SCENARIO

MPEG technology is an ever developing technology in the field of communication in which speech compression is finding its space and utility. Currently speech compression is used in satellite communications, DBS TV's, TATA SKY, dish television network which uses MPEG-4,MPEG-1,MPEG-2 and MPEG-DASH. MPEG-D is being used in audio technologies like MPEG SURROUND, SAOC (Spatial Audio Object Coding) and USAC (UNIFIED SPEECH AND AUDIO CODING)..Speech compression has dynamic adaptive streaming over HTTP (DASH) technology in the current era of communication. It is also being widely used in cellular technology that includes mobile telephony and Voice over IP. [16]

CONCLUSIONS

Speech compression is a standard for designing and compressing audio and speech signals which are transmitted to the receiver end. MPEG technology and its standards define the compression techniques and provide means for filtering noisy or undesired signals. Hence, we obtain noise-free speech signal after decompression of compressed speech signal.

ACKNOWLEDGEMENTS

We are thankful to our college and guides Mrs.Poonam Pathak and Miss.Vartika Singh respectively for providing time to time help in studying the topic and providing us a background to understand it in deeper details

REFERENCES

[1] Sinha D.,Tewhik A.H, ”Low Bit Rate Transparent Audio Compression using Adapted Wavelets”, IEEE TRANSACTIONS ON SIGNAL PROCESSING. VOL. 41. NO. 12. DECEMBER 1993

[2] Salomon D, "Video Compression: Data compression” Springer ISBN 978-1-84628-602-5,2007

[3] http://mpeg.chiariglione.org/

[4] ISO/IEC TR 23000-1:2007 – Information technology – Multimedia application format (MPEG-A) – Part 1: Purpose for multimedia application formats

[5] ISO. "ISO/IEC 23002-1:2006 – Information technology – MPEG video technologies – Part 1: Accuracy requirements for implementation of integer-output 8x8 inverse discrete cosine transform

[6] ISO. "ISO/IEC 23003-1:2007 – Information technology – MPEG audio technologies – Part 1: MPEG Surround

[7] ISO. "ISO/IEC 23004-1:2007 – Information technology – Multimedia Middleware – Part 1: Architecture

[8] ISO. "ISO/IEC 29116-1:2008 – Information technology – Supplemental media technologies – Part 1: Media streaming application format protocols

[9] http://mpeg.chiariglione.org/working_documents.htm#MPEG-V

[10] Satellite Communications,Seventh Edition 2009,Agarwal D.C,Khana Publications,p-p 368-369

[11] Davis Yan Pan,”Digital Audio Compression”, Digital Technical Journal Vol. 5 No. 2, Spring 1993

[12] K. Brandenburg and J. D. Johnston, "Second Generation Perceptual Audio Coding: The Hybrid Coder," Preprint 2937, 88th Audio Engineering Society Convention, Montreaux (1990).

[13] K. Brandenburg, J. Herre, J. D. Johnston, Y. Mahieux, and E. Schroeder, "ASPEC: Adaptive Spectral Perceptual Entropy Coding of High Quality Music Signals," Preprint 3011, 90th Audio Engineering Society Convention, Paris (1991)

[14] Pan, D “A tutorial on MPEG/audio compression” MultiMedia, IEEE (Volume:2 , Issue: 2),2002

[15] Gersho, A. “Advances in speech and audio compression” Proceedings of the IEEE Volume:82 , Issue: 6 ,August 2002

[16] M. Arjona Ramírez,M. Minami, “Technology and standards for low-bit-rate vocoding methods,” in The Handbook of Computer Networks, H. Bidgoli, Ed., New York: Wiley, 2008, vol. 2, pp. 447–467

[17] Joshi Anil Kumar,\mehta Ashish, “Audio Compression using Logarithmic Approach for PSNR enhancement”

BIOGRAPHIES

Mr. Manas Arora, is a M.TECH 2nd year student of BBDNIIT college and completed my B.TECH from SRMCEM,LKO with 77.4% (HONOURS) in the year 2012.

Ms. Neha Maurya is a M.TECH 2nd year student of BBDNIIT college and completed B.TECH from SRMCEM,LKO with 74.16% in the year 2012.

Mrs. Poonam Pathak,is presently Head of Department of Electronics & Communication Department at BBDNIIT.She has done her MTECH in DIGITAL ELECTRONICS & SYSTEMS.

Ms Vartika Singh is presently working in BBDNIIT as a senior lecturer. She has done her Bachelors and Masters Degree in Electronics and Communication Engineering from Amity University, Uttar Pradesh. Presently, she has focused her working area within the various aspects of the image processing using some suitable simulation software in the vicinity of implementation of some highly efficient and accurate algorithms for digital watermarking.Her research interests also include cooperative communications, free-space optical communications and wireless communications