A Survey Paper on Different Speech Compression Techniques

Kanawade Pramila.R[1], Prof. Gundal Shital.S[2]

[1] *M.E. Electronics, Department of Electronics Engineering, Amrutvahini College of Engineering, Sangamner, Maharashtra, India.*
[2] *HOD in Electronics Department, Department of Electronics Engineering , Amrutvahini College of Engineering, Sangamner, Maharashtra, India.*

ABSTRACT

This paper describes the different types of speech compression techniques. Speech compression can be divided into two main types such as lossless and lossy compression. This survey paper has been written with the help of different types of Waveform-based speech compression, Parametric-based speech compression, Hybrid based speech compression etc. Compression is nothing but reducing size of data with considering memory size. Speech compression means voiced signal compress for different application such as high quality database of speech signals, multimedia applications, music database and internet applications. Today speech compression is very useful in our life. The main purpose or aim of speech compression is to compress any type of audio that is transfer over the communication channel, because of the limited channel bandwidth and data storage capacity and low bit rate. The use of lossless and lossy techniques for speech compression means that reduced the numbers of bits in the original information. By the use of lossless data compression there is no loss in the original information but while using lossy data compression technique some numbers of bits are loss.

Keyword: - *Bit rate, Compression, Waveform-based speech compression, Parametric-based speech compression, Hybrid based speech compression.*

---

1. INTRODUCTION -1

Speech compression is use in the encoding system. The bit rate reduction is use in the encoding system. By the use of bit rate reduction algorithm, the minimum bits are used to compare the original information [7]. There are different speech compression techniques are present. Basically is divided in two types 1.Lossy 2.Lossless. Lossy compression means a class of data compression algorithm that allows the exact original data to be reconstructed from the exact original data to be reconstructed from the compressed data but bit rate and is better than lossless. It is compression ratio is higher than lossless compression. While lossless means output signal and input signals sounds undistinguished. Speech coder analyzed using subjective and objective analysis. Subjective is making judgments by listening output and original signal. Playing back signal and checking quality. Objective includes technical assess. Such as computing segmental signal to noise ratio (SEGSNR) between original and output signal [1]. Speech compression motivation is to remove redundancy in speech representation to reduce transmission bandwidth and storage space or memory (and apart to reduce cost). The purpose of speech compression is to reduce the number of bits required to rep- resent Speech signals (by reducing redundancy) in order to minimize the requirement for transmission bandwidth (e.g., for voice transmission over mobile channels with limited capacity) or to reduce the storage costs (e.g., for speech recording). Before we start describing speech compression coding techniques, it is important to understand how speech signal is represented in its digital form, that is, the process of speech signal digitization [4].

There are in general three basic speech compression techniques, which are waveform-based, parametric-based and hybrid coding techniques.

.

1.1 Waveform-based compression-1

Waveform-based codecs are intended to remove waveform correlation between speech samples to achieve speech compression. It aims to minimize the error between the re- constructed and the original speech waveforms [9].

It is classified as time domain and frequency domain

- Time domain: such as
  - A. PCM (Pulse code modulation)
  - B. ADPCM (Adaptive Differential PCM)
- Frequency domain or Transform coding: such as
  - A. Fast Fourier Transform (FFT)
  - B. Discrete Cosine Transform (DCT)
  - C. Continuous Wavelet Transform (CWT)
  - D. Discrete Wavelet Transform (DWT)

Waveform coders are able to produce original signal at decoder (Lossless). Bit rate range - 64 kb/s to 16 kb/s. At bit rate lower than 16 kb/s, the quantization error for waveform based speech compression coding is too high, and this results in lower speech quality [2,3].

1.1.1 Time domain

Time domain is the type of data compression for natural data like audio signal or photographic images. The remaining information can then be compressed via variety of methods. PCM, ADPCM are the types of time domain used for data transform into another mathematical domain for suitable compression [19,5].

1.1.1.1 PCM (Pulse code modulation)

The history of audio and music compression begin in the 1930s with research into pulse-code modulation (PCM) and PCM coding. Compression of digital audio was started in the 1960s by telephone companies who were concerned with the cost of transmission bandwidth. For PCM, it uses non-uniform quantization to have more fine quantization steps for small speech signal and coarse quantization steps for large speech signal (logarithmic compression) [11]. Statistics have shown that small speech signal has higher percentage in overall speech representations. Smaller quantization steps will have lower quantization error, thus better Signal-to-Noise Ratio (SNR) for PCM coding. There are two PCM codecs, namely PCM -law which is standardized for use in North America and Japan, and PCM A-law for use in Europe and the rest of the world. ITU-T G.711 was standardized by ITU-T for PCM codecs in 1988 [14]. For both PCM A-law and -law, each sample is coded using 8 bits (compressed from 16-bit linear PCM data per sample), this yields the PCM transmission rate of 64 kb/s when 8 kHz sample rate is applied (8000 samples/s 8 bits/sample =64 kb/s). 64 kb/s PCM is normally used as a reference point for all other speech compression codecs [13].

1.1.1.2 ADPCM (Adaptive Differential PCM)

Adaptive Differential Pulse Code Modulation (ADPCM), another method of speech coding, was also first conceived in the 1970s. In 1984, the United States Department of Defense produced federal standard. Differential coding refers to coding the difference between two signals rather than the signals themselves [12]. In differential coding, the short-term redundancy of the speech waveform is removed as much as possible. This is accomplished by forming an *error (difference) signal* by subtracting an estimate of the signal from the original signal. The estimate is generally obtained by a linear predictor that estimates the current samples from a linear combination of one or more past samples. The main source of performance improvement for DPCM coders is the reduced dynamic range of the quantizer input signal. Since the quantization noise is proportional to the step size, a signal with a smaller dynamic range can be coded more accurately with a given number of quantization levels. ADPCM provides greater levels of prediction gain than simple DPCM depending on the sophistication of the adaptation logic and the number of past samples used to predict the next sample. The prediction gain of ADPCM is ultimately limited by the fact that only a few past samples are used to predict the input and the adaptation logic only adapts the quantizer not the prediction weighting coefficients. ADPCM, proposed by Jayant in 1974 at Bell Labs, was developed to further compress PCM codec based on correlation between adjacent speech samples [14, 15].

1.1.2 Frequency domain or Transform coding

Transform coding is the type of data compression for natural data like audio signal or photographic images. In this, the knowledge of the application is used to choose the information to discard, thereby lowering its bandwidth [13]. The remaining information can then be compressed via variety of methods. FFT, DCT, CWT, DWT, [21] are the types of transform coding used for data transform into another mathematical domain for suitable compression [8, 17].

1.1.2.1 Fast Fourier Transform (FFT)

It is one of the methods for signal and image compression. FT decomposes a signal defined on infinite time interval into a λ-frequency component where λ can be real or complex number [22]. FT is actually a continuous form of Fourier series. FT is defined for a continuous time signal $x\ t$ as,

$$x\ f = x\ t\ .\ e{-}i\omega t\ \ .\ dt\ \infty\ {-}\infty \quad \text{------------}\ \ 1$$

The above equation is called as analysis equation. It represents the given signal in different form; as a function of frequency. The original signal is a function of time, whereas after the transformation, the same signal is represented as a function of frequency [20].

1.1.2.2 Discrete Cosine Transform (DCT)

The discrete cosine transform (DCT) helps separate the image into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality) [18]. The DCT is similar to the discrete Fourier transform: it transforms a signal or image from the spatial domain to the frequency domain. A discrete cosine transform (DCT) expresses a sequence of finitely many data points in terms of a sum of cosine functions oscillating at different frequencies. DCTs are important to numerous applications in science and engineering, from lossy compression of audio (e.g. MP3) and images (e.g. JPEG) [7].

1.1.2.3. Continuous Wavelet Transform (CWT)

The drawbacks inherent in the Fourier methods are overcome with wavelets. A wavelet is a waveform of effectively limited duration that has an average value of zero fig 2.2 wavelet is a waveform of effectively limited duration that has an average value of zero. Fourier analysis consists of breaking up a signal into sine waves of various Frequencies. Similarly, wavelet analysis is the breaking up of a signal into shifted and scaled versions of the original (or mother) wavelet [24].

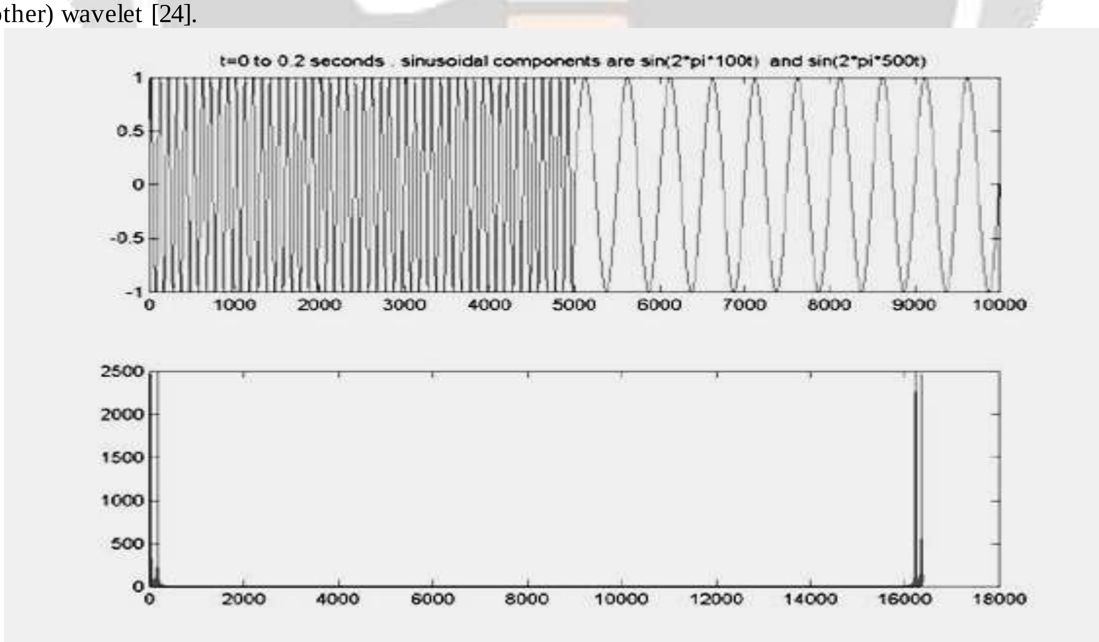

Figure 2.1: signal x2(t) and its FFT

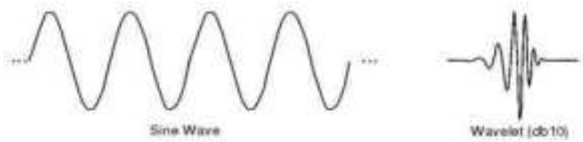

Figure 2.2: comparing sine wave and a wavelet

1.1.2.4 Discrete Wavelet Transform (DWT)

DWT is based on sub-band coding, is found to yield a fast computation of wavelet transform. It is easy to implement and reduce the computation time and resources required. In CWT, the signals are analyzed using a set of basic functions which relate to each other by simple scaling and transition. In case of DWT, the time scale representation of the digital signal is obtained using digital filtering techniques. The signal to be analyzed is passed through filters with different cut-off frequencies at different scales [6].

2. PARAMETRIC-BASED SPEECH CODING -2

Parametric-based compression methods are based on how speech is produced. Instead of transmitting speech waveform samples, parametric compression only sends relevant parameters related with speech production to the receiver side and reconstructs the speech from the speech production model. Thus, high compression ratio can be achieved. Bit rate range - 1.2 kb/s to 4.8kb/s [25].

2.1 Linear Predictive coding (LPC)-1

The history of speech coding makes no mention of LPC until the 1970s. However, the history of speech synthesis shows that the beginnings of Linear Predictive Coding occurred 40 years earlier in the late 1930s. The first vocoder was described by Homer Dudley in 1939 at Bell Laboratories [24]. Linear Predictive coding (LPC) is one of the most powerful and useful speech analysis techniques for encoding good quality speech at a low bit rate. It provides extremely accurate estimates of speech parameters, and is relatively efficient for computation [10].

3. HYBRID SPEECH COMPRESSION -3

Many different techniques are explored to represent waveform-based excitation signals such as multi-pulse excitation, codebook excitation and vector quantization. The most well known one, so called Codebook Excitation Linear Prediction (CELP)‖ has created a huge success for hybrid speech codec in the range of 4.8 kb/s to 16 kb/s for mobile/wireless/satellite communications [23].

Types of Hybrid speech compression:

A. Codebook Excitation Linear Prediction (CELP)

B. Vector Sum Excited Linear Predictive Coder (VSELP)

3.1 Codebook Excitation Linear Prediction (CELP)-1

(CELP) algorithms and Vector Selectable Excited Linear Predictive (VSELP) algorithms were developed in the mid-1980s and used commercially for audio music coding in the later part of that decade. The 1990s have seen improvements in these earlier algorithms and an increase in compression ratios at given audio quality levels [15]. CELP- Medium or low bit-rate speech coders have been researched for application to mobile radio communications. Code excited linear prediction (CELP) coding is one of the most effective coding methods at low bit-rates, which was proposed in the mid-eighties by and Schroeder and Atal [5]. CELP algorithm can produce low-rate coded speech comparable to that of medium-rate waveform coders thereby bridging the gap between waveform coders and vocoders. CELP is an efficient closed loop analysis-synthesis method for narrow and medium band speech coding systems [4-16 kbps]. In CELP coders, speech is segmented into frames (typically 10-30 ms long) and for each frame an optimum set of linear prediction and pitch filter parameters are determined and quantized. Each speech frame is further divided into a number of sub frames (typically 5 ms) and for each sub frame an excitation codebook is searched to find the input vector to the quantized predictor system that gives the best reproduction of the speech signal [26, 16].

3.2 Vector Sum Excited Linear Predictive Coder (VSELP)-2

Gerson and Jasiuk [3] proposed a vector sum excited linear predictive (VSELP) code which is associated with fast codebook search and robustness to channel errors, for use in digital cellular and mobile communications. An 8 kbps

VSELP coder was selected by the Telecommunications Industry Association (TIA) as the standard for use in North American digital cellular telephone systems. The code books in the VSELP encoders are organized with a predefined structure which significantly reduces the time required for the optimum code word search. The VSELP codec utilizes three excitation sources [21,22].

4. CONCLUSIONS

This survey paper discusses the different type of lossy and lossless data compression and different type of research paper on the based on data compression. This paper only discusses the general idea of speech compression. Today, many compression techniques are developed and some techniques are in process .But this paper only discusses the general idea about the Waveform-based speech compression, Parametric-based speech compression and Hybrid based speech compression. Parametric based codec are higher in implementation complexity but can achieve better compression ratio. This paper has been written to understand Parametric-based speech compression in the better manner and relate it to the future work.

5. ACKNOWLEDGEMENT

We would like to acknowledge Electronics Engineering Department of Amrutvahini College of Engineering and all the people who provided with the facilities being required and conductive conditions for completion of the review paper.

6. REFERENCES

[1] Saumya Mishra, Shraddha Singh *"A Survey Paper on Different Data Compression Techniques",* Indian journal Research Paper, Volume : 6, Issue : 5, May 2016.

[2] J. N. Holmes "*The JSRU Channel Vocoder*"IEEE Proc., Vol: 127 PP: 53-60, 1980.

[3] J. Lukasaik, C. McElory, E. Chang, *"Compression Transparent Low-Level Description of Audio Signals*",IEEE Int. Conf. on Multimedia and Expo, 6 July 2005.

[4] EmrahAkyol, Kenneth Rose, "*A Necessary and Sufficient Condition for Transform Optimality in Source Coding",* IEEE Int. Symp. on Information Theory Proceedings, 2011, pp. 2597-2601.

[5] TejaswiNanjundaswamy, Kenneth Rose, "*Cascaded Long Term Prediction for Coding Polyphonic Audio Signals",*IEEE w/s on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, Oct. 2011.

[6] RajeshreeRaut& Kishore Kullat, "*SDR Design with Advanced Algorithms for Cognitive Radio*" IJACS, Vol.1, No.4, pp: 134-141, Oct 2011.

[7] PramilaSrinivasan, Leah H. Jameason, "*High Quality Audio Compression Using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modelling*",IEEE Trans. On Signal Proc., vol. XX, No. Y, 1999.

[8] J. Makhoul, "*Linear Prediction: A Tutorial review*",Proc. IEEE, vol. 63(4), pp. 561-580, 1975.

[9] Daniele Giacobello, Mads G. Christensen, Manohar N. Murthi, Soren H. Jenson, Marc Moonen, "*Speech Coding Based on Sparse Linear Prediction"*IEEE Trans. Acoustics, speech, and signal processing vol. 15, no. 5, 2010.

[10] Shijo M. Joseph, Firoz Shah BabuAnto P, "*Comparing Speech Compression Using Waveform Coding and Parametric Coding",*International Journal of Electronics Engineering, 3 (1), 2011, pp. 35– 38.

[11] Daniele Giacobello, Mads G. Christensen, Manohar N. Murthi, Soren H. Jenson, Marc Moonen, " *Retrieving Sparse Patterns Using a Compressed Sensing Framework: Applications to Speech Coding Based on Sparse Linear Prediction",*IEEE Signal Proc. Letters, vol. 17, no. 1, Jan 2010.

[12] ShanonWichman, "*Comparison of speech coding algorithms ADPCM Vs CELP*,"Department of Electrical Engineering, Texas univ. 1999.

[13] Allen Gersho, "*Advances in Speech and Audio Compression*",IEEE Proc. Vol.82, No. 6,June 1994.

[14] G. Kemper, Y. Iano," *An Audio Compression Method Based on Wavelets Subband Coding*",Latin America Transactions, IEEE (Revista IEEE America), vol.09, no. 05, pp. 610-621 Sept. 2011.

[15] Stefan Strahl, Heiko Hansen, Alfred Mertins, "*A Dynamic Fine-Grain Scalable Compression Scheme With Application to Progressive Audio Coding",* Audio, Speech, and Language Processing, IEEE Transactions, vol. 19, no. 1, pp. 14-23, Jan 2011.

[16] S. Shirani, "*Multimedia communications Sub-band coding*", McMaster Univ.

[17] YiengZang, Mehmet Zeytinoglu, "*Improved Lifting Scheme for Block Subband Coding*"in the proceedings of IEEE SP. International Conference on Electronic Devices, Systems and Applications(ICEDSA1999), pp.: 487-490, 1999.

[18] Y. Zhang and M. Zeytinoglu, "*Block Subband Coding and Time-Varying Orthogonal Filter banks*"in the proceedings of 1999 IEEE PacRim'99 Conf. on Comm. , Comp. Signal Proc.

[19] Martin Vetterli, Cormac Herley, " *Wavelets and Filter Banks: Theory and Design" ,* IEEE Trns. On Signal Proc. Vol. 40, No. 9, Sept. 2009, pp. 2207-2232.

[20] R.A. Gopinath and C.S. Burrus, "*Factorization Approach to Time-Varying Unitary Filter bank Trees*"CML TR-92-23, Department of Computer and
Electrical Engineering. Rice University, Houston, USA, 1992.

[21] M. Ramya, M. Sathyamoorthy, "*Speech Coding by Using Subband Coding"*,ISBN: 978-1-4675-2248-9, ICCCE 2012, 12 and 13 Apr 2012.

[22] J. N. Holmes "*The JSRU Channel Vocoder*"IEEE Proc., Vol: 127 PP: 53-60, 1980.

[23] Samsul A.A. Karim, Bakri A. Karim, Mohd H, Kamarudin, M.K. Hasan, JumatSulaiman, "*Wavelet Transform and Fast Fourier Transform for Signal Compression: A Comparative Study,*"in the proceedings of IEEE International Conference on Electronic Devices, Systems and Applications(ICEDSA2011), pp.: 348-351, 2011.

[24] Samsul A.A. Karim, Bakri A. Ismail, Mohd H, Kamarudin, M.T. Hasan, JumatSulaiman, " *Compression of Temperature Data by Using Daubechies Wavelets*",ICMS2, 2010, pp.726-734.

[25] Mickael De Meuleneire, HerveTaddei, Olivier de Zelicourt, Dominique Pastor, Peter Jax, *"A CELP Wavelet Scalable Wideband Speech Coder*"Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference, vol. 01.