International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Special Issue 40 (KCCEMSR) (March 2016), PP. 117-120

SPEECH CODING TECHNIQUES

Kiran Chavan[1],Pranali Jawale[2],Samruddi Patil[3]
Department of Electronics and Telecommunication,
[1] Fr.C.Rodrigues Institute of Technology,
[2] Fr.C.Rodrigues Institute of Technology,
Navi Mumbai, IndiaCountry
[1] kiranchavn@gmail.com

Abstract—**Speech has been considered to be the most important component for human to communicate. Speech coding is the act of converting the speech signal in a more compressed form, which can be transmitted with less numbers of bits. It is nearly impossible to access unlimited bandwidth of a channel each time we send a signal, so there is a need to compress speech signals. Speech compression is required in the fields of long distance communication, high-quality speech storage, and message encryption having. Speech coding is the lossy type of coding, which means that the output signal does not exactly sound like input. In this paper different speech coding techniques namely Waveform Coding, Subband Coding And Linear Predictive Coding are discussed. Analysis of Differential Pulse Code Modulation is done by varying number of bits using MATLAB R2010a.**

Index Terms— **Waveform Coding, Subband Coding, Linear Predictive Coding.**

I.INTRODUCTION

Speech coding is a process of obtaining compressed form of speech signal. It allows efficient transmission of speech signal over a band limited wired or wireless channel. One of the most important applications of various speech coding techniques is security. The main objective for this is to represent the original speech using minimum number of bits so that there is no reduction in the perceptual quality. In wireless communication systems, bandwidth is the most important parameter to be taken into account and service providers are continuously met with the challenge of accommodating more number of users within a limited allocated bandwidth. To do this various speech coding techniques are used.

II.SPEECH CODING

Speech is a non-stationary signal whose properties changes with time. Speech is a band limited signal having a bandwidth of 300 Hz-3.4 kHz. Speech signal is mainly classified as voiced, unvoiced and silence sections. Voiced signals are generated by vibration of vocal cords during pronunciation of phoneme. On the contrary, Unvoiced signals are aperiodic vibration of vocal cords. Speech coding is an art of compressing and then encoding speech signals. Speech coding techniques are classified as shown in Fig1.

A. Waveform Coding:

Waveform coding is the simplest one of all the coding techniques. It is concerned with preserving the shape of analog speech signal to transmit a loyal representation of speech. Waveform coders are known to have low complexity which produce high quality of speech at the rates approximately about 16kbps. The quality of reconstructed signal degrades as data rate is reduced.

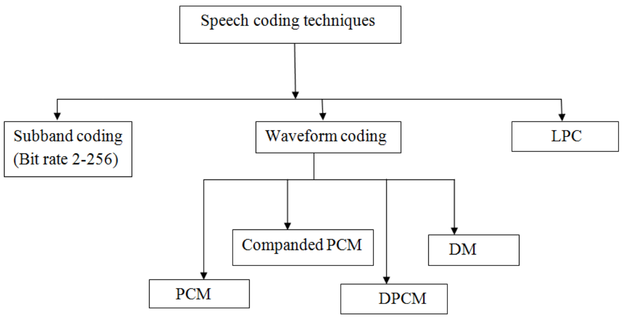

Fig.1 Classification of speech coding techniques

a. Pulse Code Modulation:

This is the method of analog to digital conversion. The operations that it involves are sampling, quantizing and coding based on quantized levels. Bit rate can be reduced by using the non-uniform quantization of the samples of speech signal. With linear quantization good quality of speech can be produced using around 12 bits per sample. Bit rate is 64 kbps.

b. Companded Pulse Code Modulation:

This method uses logarithm-based function to encode audio samples. The process of companding provides precise quantization of lower amplitude samples. In companding, at the source compression is performed and at the receiver side expansion is performed to get back the signal. Here input analog signal samples are compressed to logarithmic segments and each segment is quantized using uniform quantization. Two companding standards are: A-law and Mu-law. A-law inputs up to 13-bit samples and produces 8-bit codeword.

$$y = sgn(x)\frac{A|x|}{1+\ln(A)}$$

It reduces dynamic range compression of the signal and it increases the coding efficiency.
Mu-law inputs upto 14-bit samples and produces an output of 8-bit codewords.

$$y = sgn(x)\frac{\ln(1+\mu|x|)}{\ln(1+\mu)}$$

c. Differential *Pulse Code Modulation:*

At the time of the PCM process, the differences between input sample signals are minimal. Differential PCM (DPCM) is designed to calculate this difference and then transmit this small difference signal instead of the entire input sample signal. Since the difference between input samples is less than an entire input sample, the number of bits required for transmission is reduced. This allows for a reduction in the throughput required to transmit voice signals.

Using DPCM can reduce the bit rate of voice transmission down to 48 kbps. The first part of DPCM works exactly like PCM that is why it is called differential PCM. The sampled input signal is stored in what is called a predictor. The predictor takes the stored sample signal and sends it through a differentiator. The differentiator compares the previous sample signal with the current sample signal and sends this difference to the quantizing and coding phase of PCM, this phase can be uniform quantizing or companding with A−law or u−law. After quantizing and coding, the difference signal is transmitted to its final destination. At the receiving end of the network, everything is reversed. First the difference signal is dequantized. Then this difference signal is added to a sample signal stored in a predictor and sent to a low−pass filter that reconstructs the original input signal. DPCM is a good way to reduce the bit rate for voice transmission. However, it causes some problems that deal with degradation in voice quality. One of such problems may be due to quantization errors that occur during the approximation of samples.

DPCM quantizes and encodes the difference between a previous sample input signal and a current sample input signal. DPCM quantizes the difference signal using uniform quantization. Uniform quantization generates an SNR that is small for small input sample signals and large for large input sample signals. Therefore, the voice quality is better at higher signals. This scenario is very inefficient, since most of the signals generated by the human voice are small. Voice quality needs to focus on small signals. To solve this problem, adaptive DPCM is developed.[1]

d. Adaptive Differential Pulse Code Modulation:

ADPCM uses linear prediction that means it uses the previous samples (or several previous samples) to predict the current sample. It then calculates the difference between the current sample and its predication and quantizes the difference. For each input sample X(n), output C(n) of the encoder is simply certain number of quantization levels. Decoder multiplies this number by the quantization step to obtain the reconstructed audio sample. The method used is efficient because the quantization step is updated on the time ,by both encoder and decoder , according to the varying amplitude of the input speech samples [2]

B. *Subband Coding*:

Subband coding is the process of decomposing a speech signal into subbands using bandpass filter and coding each band separately. To reduce number of samples, sampling rate of the signal in each subband is reduced by decimation. Decimation depends on the ratio of filter output to the filter input. Decimation is also known as down sampling which follows Nyquist rule according to which range of the frequencies of output filter is more than input to the filter. As the filters are not ideal overlap between adjacent bands or aliasing occur during decimation. Filter output after decimation is quantized and encoded using different encoding schemes like PCM, ADPCM or analysis or synthesis filter. The advantage of subband coding is that each of the subband can be encoded separately using different coders based on perceptual characteristics. Decoding and upsampling of encoded samples is done at the receiver which is further passed through reconstruction filter to get back the signal.[3]

C. *Linear Predictive Coding*:

LPC is considered as a lossy type of coding because the output is not same as the input. LPC is powerful, good quality and is having a lower bit rate compared to other techniques. Two main components are there, they are LPC synthesis and LPC analysis. LPC analysis checks whether the signal is voiced or unvoiced and also finds the pitch of each signal and the parameters required to build the source filter model. At the receiver side we send the parameters obtained from the transmitter side this is done while LPC synthesis. It is an all pole filter.[4][5]

$$H(z)=\frac{G}{1+\sum a_p Z^{-p}}$$

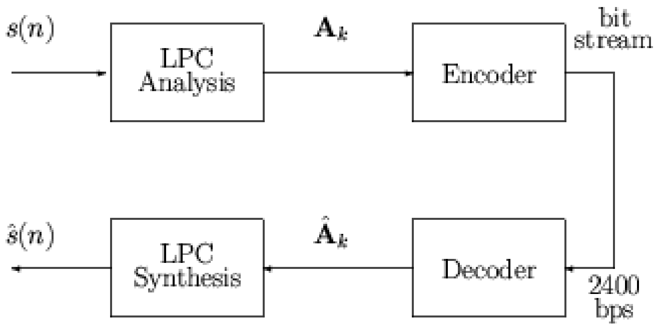

Fig.3 Block diagram of LPC

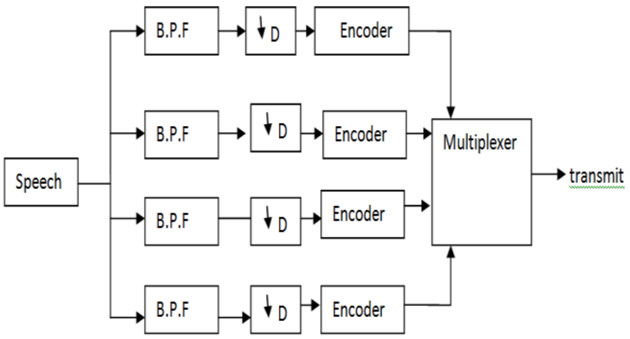

Fig.2 Block diagram of subband coding

III.IMPLEMENTATION

The speech coding starts with recording the speech sample using an AUDACITY software which will be then used as an input to the PCM coder.The steps to implement Differential Pulse Code Modulation are:

- Read the recorded speech signal in MATLABR2010a and convert it to the discrete form.
- Calculate total length and maximum amplitude of the signal.
- Input the number of bits which decides the number of quantization levels.
- Now the difference between the successive samples is calculated and are approximated to the nearest quantization level.
- Codes corresponding to each of the sample are then assigned to obtain the codeword.
- PCM representation for each sample of the input speech is available.
- Inverse of all the steps is performed In order to perform decoding.
- Observe the decoded signal and compare it with the original signal and calculate mean square error and signal to noise ratio.
- Change the number of bits used for coding and compare the changes In the obtained demodulated signal.

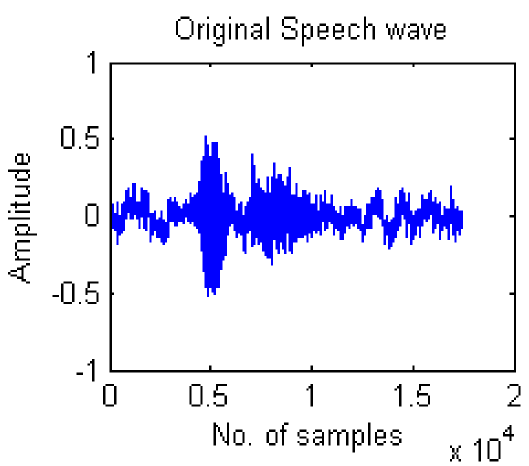

Fig.4 Original speech signal

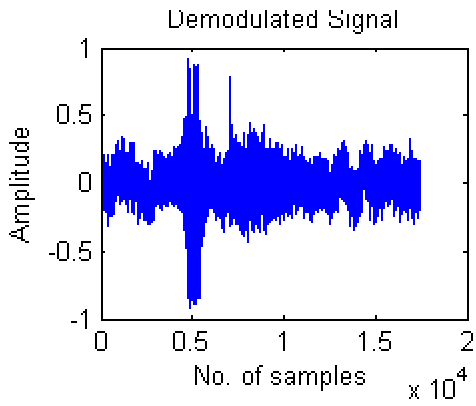

Fig.5 Reconstructed signal using 2bits per sample

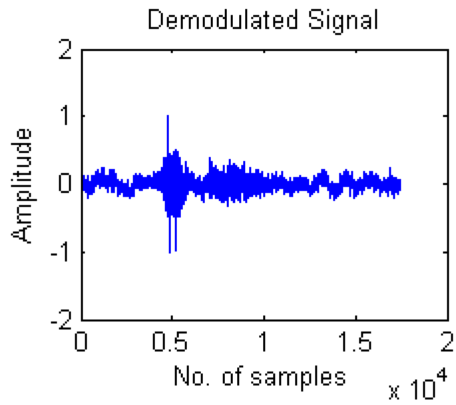

Fig.6 Reconstructed speech signal using 4bits per sample

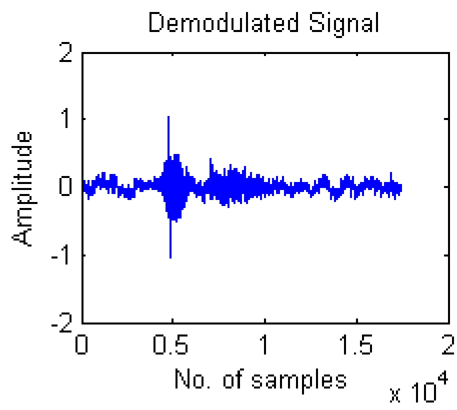

Fig.7 Reconstructed speech signal using 8bits per sample

IV.RESULTS

As the number of bits are increased, decoded signal tend to be identical to the original signal. Less number of bits used for ecoding results in degradation in the quality of the reconstructed signal.

TABLE I. MSE and SNR of signal by varying number of bits

Number of bits	Mean square error	Signal to noise ratio
2	0.0152	-10.2503
4	7.342×10^{-4}	-2.9977
8	3.679×10^{-5}	27.7232

The audience poll was taken wherein 10-15 students were asked to listen to the original sound and the reconstructed sound as well. Almost 11 students observed that when less number of bits were used the sound was not that clear as that of the original.

V. CONCLUSION

DPCM implementation is done using MATLAB R2010a. DPCM provides more compression as compared to PCM and companded PCM, as in case of DPCM, difference between successive samples is quantized. Reconstructed speech signal was judged on the basis of MSE and quality criteria. Mean Square Error reduces as number of bits are increased whereas signal to noise ratio increases.

VI. ACKNOWLEDGEMENT

We sincerely thank our guide Professor Savitha Upadhya FCRIT,Vashi, Mumbai, whose advice and suggestions provided us guidance to complete our analysis.

VII. REFERENCES

[1] David Saloman, “Data Compression: The complete reference”, Fourth edition Springer publication
[2] *Waveform Coding Techniques Document ID: 8123*
[3] Jerry D. Gibson, “Speech coding methods, standards and applications”, *University of California, Santa Barbara*, accessed in August.
[4]www.hum.uu.nl/uilots/lab/courseware/phometics/LPC , accessed in August
[5] S.K.Jagtap, M.S.mulye, M.D.Uplane, “*Speech Coding Techniques*”, *ICAC3’15, Science Direct*