Advances in Research
8(3): 1-6, 2016; Article no.AIR.30234
ISSN: 2348-0394, NLM ID: 101666096

SCIENCEDOMAIN *international*
www.sciencedomain.org

Low Bit Rate Speech Coding Using Differential Pulse Code Modulation

Shahnawaz Uddin[1], Iffat Rehman Ansari[1] and Sameena Naaz[2*]

[1]University Women's Polytechnic, Aligarh Muslim University, Aligarh, India.
[2]Department of CSE, Faculty of Engineering and Technology, Jamia Hamdard University, New Delhi, India.

Authors' contributions

This work was carried out in collaboration between all authors. Author SU studied the algorithm and simulated it, performed the statistical analysis and wrote the first draft of the manuscript. Authors IRA and SN managed the literature searches, analyses of the study/results and refined the manuscript. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AIR/2016/30234
Editor(s):
(1) Sergii Kavun, Department of Information Technologies, Kharkiv Institute of Banking, The National Bank of Ukraine (Kyiv), Ukraine.
(2) Akash Dixit, Department of Mechanical Engineering, Oakland University, USA.
Reviewers:
(1) Anonymous, University of California, Santa Barbara, USA.
(2) Anusha Yellamsetty, The University of Memphis, USA.
(3) Zhongpeng Wang, Zhejing University of Science and Technology, China.
(4) Zhao Jian, Department of Electrical and Computer Engineering, National University of Singapore, Singapore.
Complete Peer review History: http://www.sciencedomain.org/review-history/17487

Original Research Article

Received 26th October 2016
Accepted 5th January 2017
Published 10th January 2017

ABSTRACT

Generally, the voice samples are very much correlated with their past/ future samples and this feature has been exploited in the present research work. Here, various voice samples have been encoded with lower number of bits as compared to the original voice samples without much deterioration in the voice quality. We have developed a code using C-programming language on UNIX platform by applying Differential Pulse Code Modulation (DPCM) algorithm (using 4, 5, & 6-bits/ sample) on the original voice samples in Pulse Code Modulation (PCM) form (i.e., 8-bits/ sample). The results, thus obtained, differ in quality depending upon the correlation between the adjacent samples of the original voice and number of bits/ sample used for encoding.

Keywords: Speech coding; bit rate; signal to noise ratio; linear predictive coding; DPCM.

**Corresponding author: E-mail: samy.naaz@gmail.com;*

1. INTRODUCTION

Signal coding is the process of representing a signal in such a way that it realizes a desired communication objective. In this process, the signal (e.g. voice) is first sampled, quantized, and then coded so that it can be transmitted over a given channel with less probability of error under given constraints and also maximum information can be transmitted to the receiver side within minimum possible time [1]. Usually, the transmitter and the receiver are carefully designed so as to minimize the effects of noise and distortion on the quality of transmission and reception. In a PSTN (Public Switched Telephone Network), a voice channel is band-limited having a nominal bandwidth of 3.1 kHz. To estimate the data rate that can be supported by a voice channel with its limited bandwidth, Nyquist's theorem can be used which is applied to noiseless channels and it states that [2]:

$$\mathrm{R_b} = 2 * \mathrm{H} * \log_2(\mathrm{L}) \tag{1}$$

where, R_b = Maximum Data Rate in bits per second (bps), H = Channel Bandwidth, L= Number of Discrete Levels in the Signal

In case of ideal noiseless channel, the maximum data rate comes out to be 6 kbps for a 3 kHz channel and a binary signal but the maximum data rate would be decreased in case of a practical channel. In a noiseless channel, the bit rate may be increased arbitrarily by raising the number of levels used to represent the signal but as far as a noisy channel is concerned, the bit rate has an absolute maximum limit because the difference between two adjacent signal levels becomes comparable to the noise levels. For noisy channels affected by random or thermal noise, Shannon extended the Nyquist's work and Shannon's theorem states that [1]:

$$\mathrm{R_b} = \mathrm{H} * \log_2\left(1 + \frac{\mathrm{S}}{\mathrm{N}}\right) \tag{2}$$

where, R_b = Maximum Bit Rate obtainable in bits per second (bps), H = Channel Bandwidth, S/N = Signal-to-Noise Ratio (SNR).

Typical signal to thermal noise ratio of a telephone channel is about 30 dB. Therefore, the maximum bit rate comes out to be 29.9 kbps for H = 3 kHz. Hence, it can be concluded that the speech/data compression techniques are needed for the following purposes:

1. To decrease the data rate by removing the redundancy
2. To conserve the channel bandwidth
3. Minimize the transmission cost
4. Economical storage of data bits

The voice transmission/storage applications need speech coding at low rates which is known as low bit rate speech coding [3]. The overall system performance is influenced by the type of modulation used and the compression techniques. The main objective is to reduce the bit rate for ease of transmission and storage, with a non-zero but hopefully minimum possible amount of signal degradation. There are several existing waveform coders which transmit speech waveform information in the data-rate range of 16–64 kbps and the speech obtained after reconstruction has broadcast quality. In order to reduce the data-rate from the original 64 kbps, the low bit rate coders are designed in such a way that they exploit the correlation between the voice samples. At a given time instance, the speech sample may be predicted as a properly designed linear combination of a finite number of values of past sample due to the presence of this correlation. This type of prediction procedure is known as linear prediction and a number of waveform coders such as DPCM, delta modulation (DM), adaptive delta modulation (ADM), and adaptive differential pulse code modulation (ADPCM) are based on this kind of prediction. These coding schemes can be found in detail in the references [4-8]. For speech waveform coding, the ADPCM was also implemented as the G.726 ITU-T standard in 1991 [9].

Every coding/decoding method has its pros and cons. The advantages with the DPCM algorithm of voice compression are (i) removes redundancy in input signal as much as possible, (ii) quality of the signal is not deteriorated much, (iii) very simple to implement, (iv) less memory requirements, & (v) very much economical. Because of these reasons, we were motivated to choose DPCM technique to compress the correlated voice signals in 4-bits/ 5-bits/ 6-bits per sample and obtained very good compression ratio, good SNR, & fair quality of the signal as expected.

2. PERFORMANCE MEASURES

To minimize the bit rate of the digitally represented signal while retaining required level of signal quality at the same time, the general problems with voice compression technique are complexity in implementation and communication delay. The major parameters used to measure its performance are signal quality, bit rate, complexity and communication delay.

2.1 Signal Quality

In order to test the speech quality, the perceived signal quality is often measured on a five - point scale that is termed as the Mean Opinion Score (MOS), which is an average of various speech inputs, speakers, and hearers or listeners evaluating signal quality. The five points of quality are associated with a set of standardized objective description: bad, poor, fair, good and excellent. In the field of speech coding, MOS evaluations are well accepted and sometimes supplemented with measurements of intelligibility and acceptability.

2.2 Bit Rate

The bit rate of the digital representation is measured in bits per sample or bits per seconds. The rate in bits per seconds is merely the product of the sampling rate and the number of bits per sample. Typical sampling rates are 8 kHz for telephone speech, 16 kHz for AM-Radio grade audio, and 44.1 kHz or 48 kHz for CD (Compact Disc) audio / DAT (Digital Audio Tape) audio.

2.3 Complexity

In signal processing hardware, a number of computational efforts are needed for the implementation of the encoding as well as decoding process and this is known as the complexity of a coding algorithm. To measure the complexity, two important parameters are used namely, arithmetic capability and memory requirement. Coding algorithms of significant complexity are currently being implemented in real time. Other related measures of coding complexity are the physical size of the encoder or codec (encoder + decoder), their cost (in dollars) and the power consumption (in watts or milliwatts).

2.4 Communication Delay

Increasing complexity in a coding algorithm is usually associated with increased processing delays in the encoder and the decoder but the need is to keep the communication delays as minimum as possible.

3. PERFORMANCE COMPARISON OF DIFFERENT PULSE CODING METHODS

Table 1 shows the performance comparison of PCM, DPCM, Delta Modulation (DM) and Adaptive Delta Modulation (ADM). In this table, various parameters such as number of transmitted bits per sample, quantization error, transmission bandwidth, SNR, etc. are compared [10].

Table 1. Comparison between PCM, DM, ADM and DPCM

S. no.	Parameter	PCM	DM	ADM	DPCM
1.	No. of bits	It can use 4, 8, or 16 bits/ sample	It uses only 1-bit/sample	Only 1-bit/ sample	Bits can be more than 1-bit/sample but less than PCM
2.	No. of Levels/Step Size	No. of levels depend on no. of bits/fixed	2/fixed and can't vary	2/According to the signal variation	No. of levels depend on no. of bits/fixed
3.	Quantization error and Distortion	Depends on number of levels	Slope overload distortion & granular noise is present	Quantization error is present but other errors are absent	Slope overload distortion and quantization noise is present
4.	Transmission channel bandwidth and SNR	Highest as no. of bits are high	Lowest	Lowest	Lower than PCM
5.	Feedback	There is no feedback in transmitter or receiver.	Feedback exists in transmitter	Feedback exists	Feedback exists
6.	Area of Applications	Audio and Video Telephony	Speech and Images	Speech and Images	Speech and Video

S. no.	Parameter	PCM	DM	ADM	DPCM
Comparison for voice encoding only					
7.	Sampling rate (kHz)	8	64-128	48-64	8
8.	No of Bits/Sample	7-8	1	1	4-6
9.	Bit Rate (kbps)	56-64	64-128	46-64	32-48

4. IMPLEMENTATION OF DPCM

The algorithm for implementation of DPCM at the transmitter and the receiver end is as given below.

For Transmitter:

Step 1: Initialize the predicted sample value to zero.
Step 2: Get the first sample value from the source file (PCM data file with .au extension).
Step 3: Evaluate the difference between source & prediction sample and store it into e[n] that is error signal.
Step 4: Quantize this evaluated error signal using uniform midriser type quantizer.
Step 5: Encode and transmit the quantized error signal.
Step 6: Obtain the next prediction by adding previous prediction and present quantized error.
Step 7: Repeat steps 2 to 6 till the end of file.

Table 2. DPCM algorithm at the transmitter side

Input:
- Instance $x \in X$; Get the first sampled value out of N-samples from the source voice file in PCM format; x[0]
- Initial Predicted Value; $0 \leftarrow \hat{x}[0]$
- **while** not termination condition **do**
- Calculate error signal; e[n]$\leftarrow x[n] - \hat{x}[n]$
- Quantize e[n]; $q_e[n] \leftarrow \frac{del}{2} + del * floor(e[n])$
- Encode q_e[n]; $b[n] \leftarrow binary(q_e[n])$
- Next Predicted Value; $\hat{x}[n] \leftarrow \hat{x}[n-1] + q_e[n-1])$
- **end of source file**

Output:
- N-binary codes generated; b[n]

For Receiver:

Step 1: Initialize the received signal equal to zero.
Step 2: Receive the coded error signal and decode it.
Step 3: Obtain the final received signal by adding or subtracting the present error signal to the previously received signal depending upon the nature of error.
Step 4: Repeat step 2 and 3 till the end of file.

Table 3. DPCM algorithm at the receiver side

Input:
- Initial Received Signal value; $0 \leftarrow r[0]$
- **while** not termination condition **do**
- Decode received error signal; $r_e[n] \leftarrow decode(r_e[n])$
- Next Received Signal Value; $r[n] \leftarrow r[n-1] + r_e[n]$
- **end of received file**

Output:
- N-Received Signal values generated; r[n]

5. RESULTS AND DISCUSSION

Sampling Frequency = 8 KHz

Bit Rate = Sampling Rate x No. of Bits per Sample

$$\%\ \text{Compression} = \frac{(\text{Base Bit Rate} - \text{Required Bit Rate})}{\text{Base Bit Rate}} \text{x}100 \quad (3)$$

The purpose of calculating the signal to noise ratio (SNR) is to compare the noise & signal at the same point of the communication system to ensure that the noise is not excessive. The SNR is basically the ratio of two powers, i.e., it is ratio of the signal power to the noise power. Since the powers of the signal and noise are proportioned to the amplitudes of the signal & noise. Thus,

$$\text{SNR} = \frac{P_s}{P_n} = \frac{X_s^2}{X_n^2} \quad (4)$$

In this research paper, the signal to noise ratio is calculated by the following formula:

$$\text{SNR(dB)} = 10 \log_{10}\left(\sum_{i=1}^{N} \frac{x_i^2}{(x_i - r_i)^2}\right) \quad (5)$$

where, x_i is the input signal and, r_i is the output/received signal. Here, we take sum of squared values of received signals and the error signals $(x_i - r_i)$ are squared & summed. The ratio of these sums is known as SNR.

Table 4. Samples for SNR calculation

Input	DPCM output for different no. of bits		
	4-bits	5-bits	6-bits
96.00	92.00	97.33	96.00
93.00	89.33	94.66	93.33
92.00	88.66	94.00	92.66
90.00	86.66	92.00	91.33
91.00	86.66	92.00	91.33
91.00	88.00	92.00	91.33
93.00	88.67	93.33	93.33
94.00	89.33	92.00	94.00
95.00	90.00	94.67	94.67
96.00	90.00	95.33	95.33
95.00	89.33	95.33	95.33
94.00	99.33	94.67	94.67
92.00	87.33	94.67	94.67
92.00	87.33	93.67	93.33
SNR	**25.30 dB**	**35.59 dB**	**38.68 dB**

Table 5. Observation table of different input files and their no. of Bits, SNR and MOS

Input PCM file name and no. of bits	MOS for input file	Output file name and no. of bits	Percent compression, bit rate	SNR for output file	MOS for output file
Sh0.i(8)	3.5	Sh04.0.0ut(4)	50.0%, 32 kHz	**25.30**	**2.50**
		Sh05.0.0ut(5)	37.5%, 40 kHz	**35.53**	**3.00**
		Sh06.0.0ut(6)	25.0%, 48 kHz	**38.68**	**3.50**
Sh3.i(8)	3.5	Sh34.0.0ut(4)	50.0%, 32 kHz	**12.41**	**2.00**
		Sh35.0.0ut(5)	37.5%, 40 kHz	**18.20**	**2.80**
		Sh36.0.0ut(6)	25.0%, 48 kHz	**22.30**	**3.00**
Sh4.i(8)	3.5	Sh44.0.0ut(4)	50.0%, 32 kHz	**15.49**	**2.25**
		Sh45.0.0ut(5)	37.5%, 40 kHz	**19.73**	**2.80**
		Sh46.0.0ut(6)	25.0%, 48 kHz	**25.16**	**3.20**

It is apparent from the results shown in Table 4 and Table 5 that as the number of bits for encoding the error signal increases, the signal to noise ratio also increases because of decrease of step size and correspondingly the mean opinion score (MOS) increases. These MOS scores were obtained by performing listening tests with human listeners. The basic limitation of DPCM system is that we assumed first prediction equal to be zero (in ideal case) and it takes time to track the original signal so there is always some error present in the DPCM technique and hence received signal is deteriorated from its original value. Also other limitation of DPCM is that it is impossible to transmit an uncorrelated signal in an error free manner because if the signal varies randomly then it takes more time to track the original signal. Furthermore, in case of low SNR speech signal, we can't apply DPCM instead ADPCM technique can be applied.

6. CONCLUSION

The results of DPCM coding algorithm using 4, 5 and 6-bits per sample are shown in Table 5 where various input files were processed and corresponding output voice files are generated. From the results of these files, it can be concluded that the results are very attractive provided that the file samples are correlated. But if there are files in which there is lack of correlation among the successive samples, it has been found that there are some errors and the receiver takes some time to track the original value because initial value of the predicted sample is 0. Thus, if the initial samples of the input file are large then the receiver takes some time to track the actual input and the large error occurs in some initial samples. As long as input samples do not change randomly, there is good tracking. Hence, the compression of the voice signal is improved in case of DPCM as compared to PCM but at the cost of complexity of computing algorithm and circuitry. As the number of bits per sample is decreased it has been found that there is some degradation in the signal quality. The overall conclusion is that if the signal samples are correlated then the advantages of low bit rate coding can be exploited in a very efficient manner without much degrading the quality of the voice.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Simon Haykins. Digital communications. Wiley; 2006.
ISBN-10: 8126508248,
ISBN-13: 978-8126508242
2. John Dunlop, Geoffrey Smith D. Telecommunications engineering. CRC Press; 1998.
ISBN: 0-7487-4044-9
3. Alan V. McCree. Low-bit-rate speech coding. Springer Handbook of Speech Processing. 2008;331-350.
DOI: 10.1007/978-3-540-49127-9_16
ISBN: 978-3-540-49125-5
ISBN: 978-3-540-49127-9
4. Spanias AS. Speech coding: A tutorial review. Proceedings of the IEEE. 1994;82: 1541–1582.
5. Cummiskey P. Adaptive quantization in differential PCM coding of speech. Bell Systems Technical Journal. 1973; 52:1105.
6. Gibson J. Adaptive prediction in speech differential encoding systems. Proceedings of the IEEE. 1974;68:1789–1797.
7. Gibson J. Backward adaptive prediction in multi-tree speech coders in advances in speech coding: B. Atal, V. Cuperman and A. Gersho. 1990;5–12.
8. Carl Kritzinger. Low-bit-rate speech coding. Masters' thesis in engineering sciences, University of Stellenbosch, South Africa; 2006.
9. ITU-T Recommendation G726. 40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM). ITU-T Rec.G726 Document Number E 1951, International Telecommunications Union, Geneva, Switzerland; 1991.
10. Chitode JS. Digital communication. 1st Edition, Technical Publication, Pune. 2007-2008;2.21-2.22.
ISBN: 9788184312775

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/17487