Codebook Based Digital Speech Compression

Nazish Nawaz Hussaini and Asadullah Shah
Isra University
Hyderabad, Pakistan

Abstract: *Speech compression has been an area of interest for many researchers, and their objectives remained to reduce the bit rates, also to maintain the speech quality. Adaptive Differential Pulse Code Modulation (ADPCM) was first such scheme that evolved out of PCM and in low bit rate source coding techniques; Code Excited Linear Prediction (CELP) was adopted as standard in the Department of Defense (DoD), USA. In this research a novel approach is developed and compression up to 50% is achieved with negligible loss of quality. This coding technique is developed to use a PCM coded speech, initially, and then substituting 3 to 5 Least Significant Bits (LSB) of every byte at the decoding end from a fixed codebook, randomly.*

Keywords: *PCM,, Fixed Codebook, Compression*

1. INTRODUCTION

1.1. Speech Coding

In traditional telephone networks, the standard method for converting analog voice signals to digital form, is to sample the signals at the rate of 8000 times per second and then encode each sample as an 8-bit binary representation, specified by ITU (International Telecommunication Union). The result is 64-kb/s digital data stream known in telephony as the lowest level within the digital signal hierarchy. Natural speech has a great deal of redundancy, so dropping a few segments of uncompressed speech may not affect the perceived quality very much.

1.2. Motivation for Speech Coding

Emerging applications in rapidly developing digital telecommunication networks require low bit, reliable, high quality speech coders. The need ***to save bandwidth***s, and ***to conserve memory*** in voice storage systems are two of the many reasons for the very high activity in speech coding research, and development.

1.3. Basic Coding Techniques

1) *Pulse Code Modulation:* In pulse code modulation (PCM), coding the speech signal is represented as a series of quantized values which correspond to the amplitudes of the speech samples. In uniform 128 kb/s PCM, sampling at 8 kHz and quantized with 16 bits per sample. Uniform PCM does not exploit any specific properties of speech [3].

2) *Differential Pulse Code Modulation*: This coding scheme belongs to the category of predictive techniques. Due to the fact that the speech signal is a highly correlated signal and it is often efficient to encode the difference between two consecutive samples. A coder that quantizes the difference waveform rather than the original waveform is called Differential Pulse Code Modulation. There are several variations of Differential Pulse Code Modulation. Since differences between samples are expected to be smaller than the actual sampled amplitudes, fewer bits are required to represent the differences [5].

3) *Adaptive Differential Pulse Code Modulation:* This coding scheme also belongs to the category of predictive techniques. In Adaptive DPCM (ADPCM) sample to sample correlations are exploited, achieving compression objectives, the linear predictor or/and the quantization levels are varied based on the characteristics of the past reconstructed speech signals. The same predictor/quantizer modifications are performed by the encoder and the decoder. It is a family of speech compression and decompression algorithms. A common implementation takes 16-bit linear PCM samples and converts them to 4-bit samples, yielding a compression rate of 4:1[3].

4) *Code Excited Linear Prediction:* Like all vector quantization techniques, CELP coding breaks a sampled input signal into blocks of samples, that are processed as one unit. CELP is a lossy compression algorithm used for low bit rate, operates at 4.8 kbits/sec rate [2].

2. THE NOVEL APPROACH

The technique on the proposed investigation uses speech file with .pcm format at the encoding end. Initially a fixed codebook is created to represent the 2^3=8 bit patterns. In the next step a decoder is substituting the 3 lsb of every byte randomly with one of the bit patterns of the codebook as shown in figure-1.

The technique is further implemented on 4 and 5 lsb of the speech file. This novel approach is basically intended to reduce the bit rate from at-least 33% to 50% and almost no loss of subjective quality of speech. It is assumed that this strategy fulfills the bandwidth requirements. Even the rates of speech compression techniques that exploited waveform as well as source coding methods still got lower rates, but suffered quality of speech and are not accepted as standards. It is expected

that most carefully selected bit patterns from the Codebook may provide a substitute same as before substitution of 3 bits at the encoded end. But even, if an identical pattern is not selected the expected error in each byte may be just 1 bit and we believe that this 1 bit in each byte a worst scenario may not affect or deteriorate the quality and the cost of compression ratio.

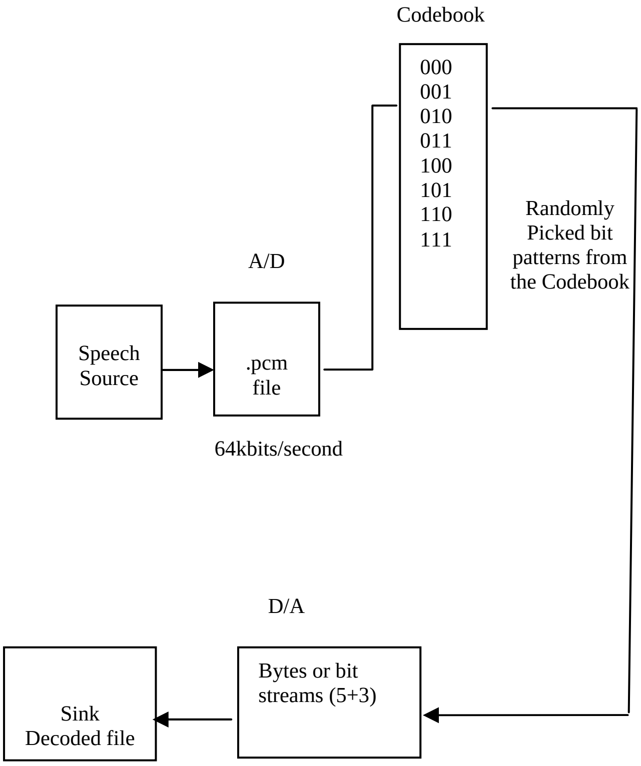


Figure 1 Schematic Diagram Showing Digital Speech Compression

3. OBJECTIVE AND SCOPE OF THIS RESEARCH

In this research many different experimentation have been carried out. Objective is to compress the speech file without any complex algorithms. The substitution of few of the bits of each byte of that file is done by a program then both the original and the substituted files are checked and viewed in binary Editor Software. For this first of all a PCM coded speech at 64 Kb/s is taken, then a logic programmed that randomly substitutes 3, 4, and 5 Least Significant Bits out of 8 bits at the decoder end from a fixed codebook of 2^3,2^4,2^5, bit patterns accordingly.

It is expected that the pattern that has been chosen for re-substitution has subjective quality loss and treated as the best scheme.

Finally the Mean Opinion Score (MOS) tests on each one of the experimentation is used to determine the quality.

This approach is expected to be a standard in the world of compression. Also if this work will be taken for saving bandwidth requirements the field of networking can also progress further.

4. LITERATURE REVIEW

[1] W. Kinser, discussed in his research paper about the success of multimedia which depends on the solution to some technical problems regarding multimedia material and I want to mention his ideas because the proposed work also need to follow these instructions (here multimedia material is sound and from those problems two are concerned with this research work): (i) how to encode? (ii) how to store and transmit its electronic representation? (iii) how to evaluate subjective quality of reconstruction ?

Speech can be compressed with respect to its dynamic range and/or spectrum. Dynamic range is the range of the lowest to the highest level that can be reproduced by a system. Digital audio at 16-bit resolution has a theoretical dynamic range of 96 dB. The dynamic range reduction is used in telephones with logarithmic A-law (Europe) or u-law (North American) capable of reducing the range from 12 bits to 8 bits only. So the uncompressed rate is 96 kbps, and the compressed rate is 64kbps.

[2] Javier and Alejandro in 2001 presented their research paper on voice compression systems for wireless telephony. Their research was based on comparison of four compression systems: CELP, VSELP (Vector Sum Excited Linear Predictor), GSM 06.10 (Global System for Mobile Communications) and ANN (Artificial Neural Networks), focusing on the fact of voiced and unvoiced sounds; they concluded that voice compression technology has not advanced up till 2001. The main cause is the speaker dependency that hinders the standardization of new voice compression systems. CELP was created in 80's and since then the research has been dedicated to its improvement, not to the establishment of a new model that achieves a substantial improvement in the performance of voice compression for wireless telephony.

[3] The digital audio data consists of a sequence of binary values representing the number of quantizer levels for each audio sample. The method of representing each sample with an independent word is called Pulse Code Modulation (PCM). Digital audio compression allows the efficient storage and transmission of data. Different audio compression techniques have different levels of complexities, quality, and amount of compressed data. Davis has surveyed digital audio compression techniques to provide experience with digital audio processing. Davis Yen Pen in 1993 discussed about process of digitization and two simple audio approaches U-law and Adaptive Differential Pulse Code Modulation. The digitization

begins with a conversion of analog to the digital domain by sampling the audio input and quantizing each sampled values into a discrete number. The number of bits/sample used for digital audio ranges from 8 to 16.
[4] Srivatsan Kandadai and Charles D. Creusere in 2006, proposed their study on Perceptually-Weighted Audio Coding that Scales to Extremely low Bit rates. A perceptually scalable audio coder generated a bit stream that contains layers of audio fidelity and is encoded in such a way that adding one of these layers enhances the reconstructed audio by an amount that is just noticeable by the listener.

5. METHODOLOGY

5.1. Audio recording

The speech file used is collected from Radio Pakistan, stored on CD with .cda extension, which is then converted to .wav file by a software named "blaze media pro converter". .wav is the uncompressed audio format usually stored on Windows .

5.2. Encoding

Recorded Analog speech file in .wav format is further digitized in by adobe audition and stored at various sampling rates like 44.1 kHz, 22.05 kHz, 11.025 kHz, and 8 kHz and quantized with 16 bits and 8 bits. The research proceeds with the saving of .wav file with a sampling rate of 8kHz and a sampling size of 8 bits in a .pcm format which is raw data having no header. As this research work is based on the speech compression of bits so there was a need to see the speech file in its binary format.

5.3. Retrieval of Speech Files in 010 Editor

A Software named "010 Editor" displays binary representation of speech file. 010 Editor is a professional hex editor designed to quickly and easily edit the contents of binary files and allows viewing and editing the individual bytes of binary files. The binary representation of the speech file before and after substitution of bit pattern help in manual checking of the encoded and later on decoded file .

5.4. Fixed Codebook

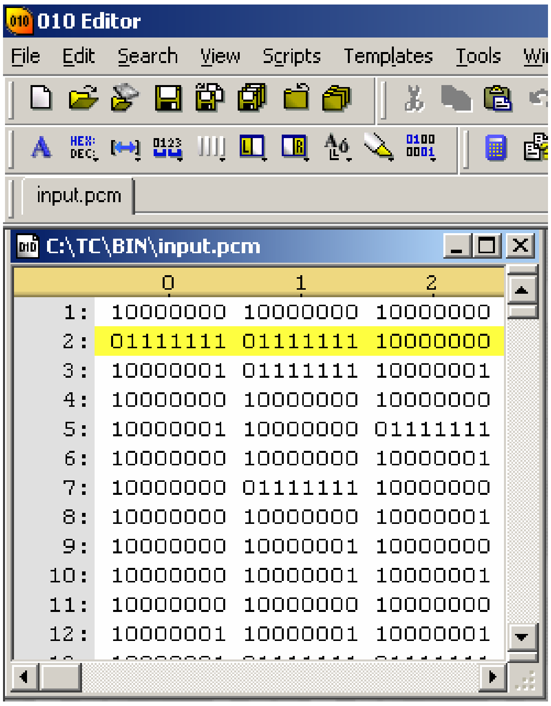
A fixed codebook is created to store 8 different bit patterns for the 3 bits i.e. 2^3=8 bit patterns, 16 different bit patterns for the 4 bits i.e. 2^4=16 bit patterns, 32 different bit patterns for the 5 bits i.e. 2^5=32 bit patterns.

5.5. Decoding

Firstly, at the decoder end, randomly pick a 3 bit pattern out of the 8 bit patterns from the codebook and substituted with the 3 Least Significant Bits of every byte of the original .pcm file. The binary representation of the input .pcm file and output file is then checked and compared as shown in Figure-2 and Figure-3.

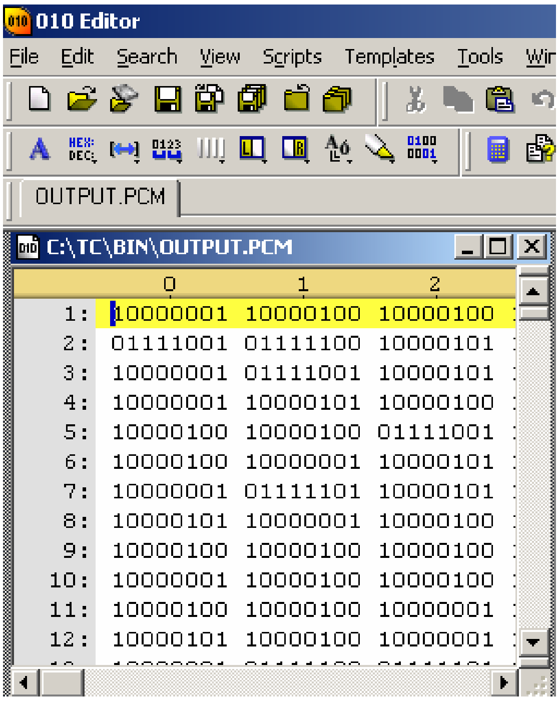
Secondly, at the decoder end, randomly pick 4 bit pattern out of 16 bit patterns from the codebook and substitute with the 4 Least Significant Bits of every byte of the original .pcm file. The binary representation of the input .pcm file and output file is then checked and compared as shown in Figure-2 and Figure-4.

Finally, at the decoder end, randomly pick a 5 bit pattern out of 32 bit patterns from the codebook and substitute with the 5 Least Significant Bits of every byte of the original .pcm file. The binary representation of the input .pcm file and output file is then checked and compared as shown in Figure-2 and Figure-5.



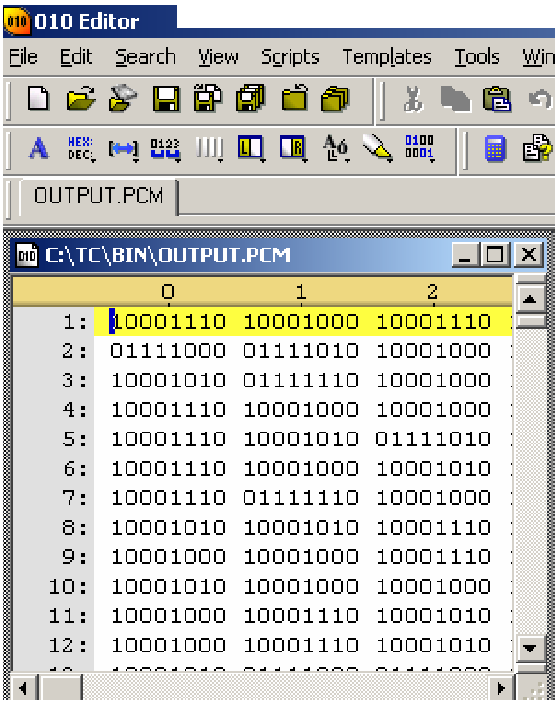
```
           0        1        2
  1:  10000000 10000000 10000000
  2:  01111111 01111111 10000000
  3:  10000001 01111111 10000001
  4:  10000000 10000000 10000000
  5:  10000001 10000000 01111111
  6:  10000000 10000000 10000001
  7:  10000000 01111111 10000000
  8:  10000000 10000000 10000001
  9:  10000000 10000001 10000000
 10:  10000000 10000001 10000001
 11:  10000000 10000000 10000000
 12:  10000001 10000001 10000001
```

Figure 2. Original input file with .pcm



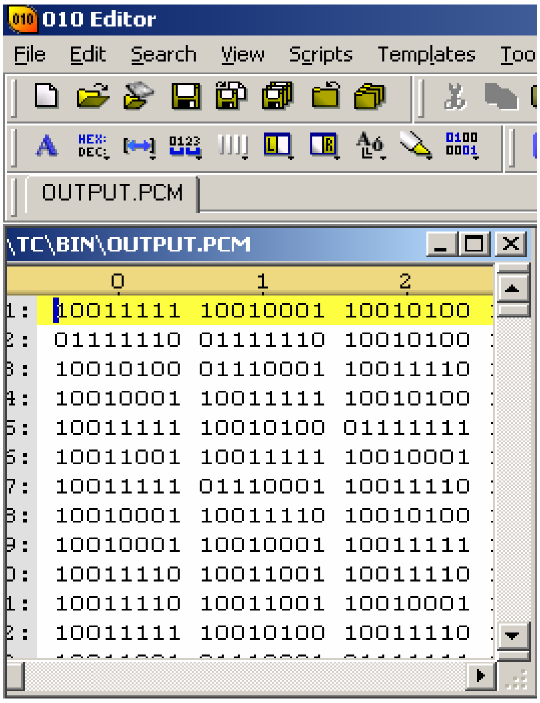
```
            0        1        2
 1:  10000001 10000100 10000100
 2:  01111001 01111100 10000101
 3:  10000001 01111001 10000101
 4:  10000001 10000101 10000100
 5:  10000100 10000100 01111001
 6:  10000100 10000001 10000101
 7:  10000001 01111101 10000101
 8:  10000101 10000001 10000100
 9:  10000100 10000100 10000100
10:  10000001 10000100 10000100
11:  10000100 10000100 10000001
12:  10000101 10000100 10000001
```

Figure 3. 3-bit substitution out put



```
            0        1        2
 1:  10001110 10001000 10001110
 2:  01111000 01111010 10001000
 3:  10001010 01111110 10001000
 4:  10001110 10001000 10001000
 5:  10001110 10001010 01111010
 6:  10001110 10001000 10001010
 7:  10001110 01111110 10001000
 8:  10001010 10001010 10001110
 9:  10001000 10001000 10001110
10:  10001010 10001000 10001000
11:  10001000 10001110 10001010
12:  10001000 10001110 10001010
```

Figure 4. 4-bit substitution out put



```
         0        1        2
1:  10011111 10010001 10010100
2:  01111110 01111110 10010100
3:  10010100 01110001 10011110
4:  10010001 10011111 10010100
5:  10011111 10010100 01111111
6:  10011001 10011111 10010001
7:  10011111 01110001 10011110
8:  10010001 10011110 10010100
9:  10010001 10010001 10011111
0:  10011110 10011001 10011110
1:  10011110 10011001 10010001
2:  10011111 10010100 10011110
```

Figure 5. 5-bit substitution out put

6. SUBJECTIVE QUALITY TEST (MEAN OPINION SCORE)

A subjective measure, like taking the mean of the opinions of a group of listeners and observed what the voice quality is? And how can it be quantified in reproducible manner? With some answers in hand, it is possible to pin down the factors that influence the perception of voice quality , and the steps that can be taken to ensure that an acceptable level of quality is met. The Mean Opinion Score (MOS) is a formal subjective quality evaluation scheme that involves a category judgment technique based on a selected listening audience evaluating the difference between original and the decoded file, on an integer scale from 1 (very annoying) to 5 (imperceptible).

Table 1. Mean Opinion Score Scale

Mean Opinion Score	Impairment Scale
5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very Annoying

For the Mean Opinion Score 16 male and female listeners gave their judgment and according to them the original PCM file and the files after substitution of 3, 4, and 5 least significant bits the results are shown in Figure-6.

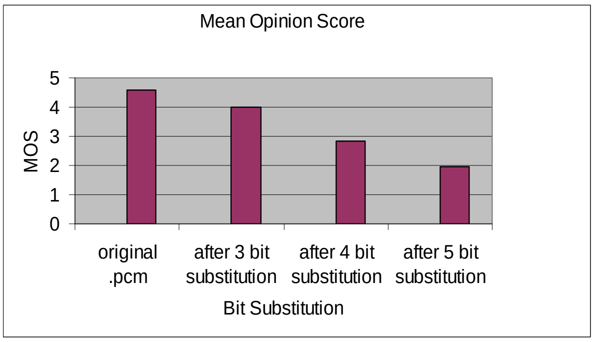


Figure 6. MOS for 4 speech files

The original .pcm file before substitution of any bit pattern but at lower sampling rate and sampling size i.e. 8K and 8bits, remarked by the listeners as perceptible and scaled at 4.59 MOS.

In the first experiment 3 bit pattern substitution in the original file does not affect understanding of speech and remarked as perceptible, but not annoying by the listeners and scaled at 4.0 MOS.

In the second experiment 4 bit pattern substitution in the original file affect understanding of speech and remarked as slightly annoying, by the listeners and scaled at 2.97 MOS.

In the last experiment 5 bit pattern substitution in the original file does not affect understanding of speech and remarked as annoying, by the listeners and scaled at 2 MOS.

7. CONCLUSION

The design of the proposed algorithm was based upon the fact that in each byte "speech sample" 3 bits can have 8 variations and from those, if selected 3 of them and randomly substituted at the decoded end shows results between very good and good quality. Further on selecting 4 bits from 16 different variations, if 4 different patterns are selected and substituted randomly at the decoded end degrades the quality. Lastly on selecting 5 bits from 32 different variations, if 5 different patterns are selected and substituted randomly at the decoded end degrades the quality more. The compression ratio for 3, 4, 5 bits is 37.5%, 50%, and 62.5%.

More careful design of the criterion used in selecting the codebook entries can be implemented to improve the results. In our tests we experienced an excellent performance of the coder/decoder.

REFERENCES

[1] W. Kinser , "Compression and its Metrics for Multimedia", Proceedings of the First IEEE International Conference on Cognitive Informatics (ICCI'02).

[2] Javeir Bustos and Alejandro Bassi, "Voice Compression Systems for Wireless Telephony", XXI International Conference of the Chilean Computer Science Society(SCCC'01), pp. 0041, November 2001.

[3] Devis Yen Pan, "Digital Audio Compression", Digital Technical Journal vol. 5 No. 2, Spring 1993.

[4] Srivatsan Kandadai and Charles D. Creusere, "Perceptually -Weighted Audio Coding that Scales to Extremely Low Bitrates", Proceedings of the IEEE Data Compression Conference (DCC'06)

[5] Corneliu Burileanu, Radu Preda, Andrei Fecioru, Dragos Ion, "Speech Compression Techniques on Motorola DSP based Platforms", Faculty of Electronics andTelecommunications, Politehnica University, Romania, Inventics Review, vol. II, 2002.