Byte Error Analysis Based on Cascaded Approach for Digital Speech Compression

Nazish Nawaz Hussaini*, Asadullah Shah**, Abdul Wahab Ansari***

Abstract : The research in the field of speech compression is still in progress, resulting the development of many compression techniques. The objective is focused on progress of speech quality. A novel cascaded approach is designed to achieve these objectives. The coding scheme works on dropping a number of least significant bits from all bytes of the pcm coded speech file at encoding end. These dropped bits at the decoding end are then submitted from the respective fixed codebooks randomly. Mean Opinion Score comparisons for analysis of error free and error occurring bytes exploits the novel strategy. The codebooks used for random substitutions of bits vary with their sizes to check the improvements in MOS.

Keywords: Compression, Fixed codebook, Mean opinion Score, cascaded approach

I. INTRODUCTION

In the field of information technology, multimedia plays a very important role. Multimedia works on multiple media like continuous and discrete media. As continuous media conserve more memory than discrete media, it needs to get compress. From all data compression schemes, speech compression is a big region to explore and deal with. Speech stored in computer is in its digital form. The need to save bandwidth in wireless and wire line communication networks and to conserve memory in voice storage systems is the main reason to explore speech coding research and development.

One of the basic waveform speech coding scheme is Pulse Code modulation (PCM). PCM encoding represents digital signals sampled uniformly and quantized as a series of numeric codes. It has been used in digital telephone systems and computers as standard form of digital audio. In digital telephony, the audio signal is encoded as 8000 samples/s, and stored on 8 bits each, i.e., 64 Kbit/s digital signal and by default that signal encoding could be either µ-law PCM or A-law PCM, described in international standard G.711[1].

*Institute of Mathematics & Computer Science, Sindh University Jamshoro , Hyderabad, 76080, Pakistan, Email:
*nazish_hussaini@hotmail.com, nazish.nawaz@usindh.edu.pk
**Department of Computer Science, IBA Sukker, dr_asadullahshah@hotmail.com
*** Department of Computer Science, Isra University
Hala Naka Road, Hyderabad,Pakistan, wahabansari@hotmail.com

Another variation of this coding scheme is Differntial PCM (DPCM). DPCM is designed to calculate the difference of the signal and transmit it. The difference of input sample is less than the entire input sample, so the number of bits required for transmission is reduced. The throughput required to transmit voice signals is 48 kbps by this scheme [2].
Code Excited Linear Prediction (CELP) is currently providing significantly better quality of speech than other low bit-rate algorithms. It is the most widely used speech-coding algorithm nowadays [3]. CELP uses codebook based strategy and for CELP to work well, the codebook must be big, and the big codebook if used then it will consume more time while searching, and require large codes. The biggest problem is that such a system would require a different code for every voice, and it makes the codebook extremely large. This problem can be resolved by using two small codebooks instead of one, i.e., fixed (codebook) and adaptive (codebook). CELP coding scheme is described in federal standard 1016, providing good quality, at 4.8 Kbits per second, some of its variants are like G. 723.1 CELP based, at 5.3 and 6.4 Kbits/sec and G.729 CELP based, at 8 Kbps [4].

Some of the basic speech compression schemes along with their data rates and MOS are collected and shown in the Table-1[5].

Table-1 Data rates and MOS of Basic compression schemes

Compression scheme/year	ITU-T	Data rate (kbps)	MOS
PCM, 1998	6.711	64	4.4
ADPCM, 1991	6.721, G.723	32 64	4.2 4.2
LDCELP	G.723.1 G.728	5.3, 6.4 16	4.2 4.2

II. LITERATURE REVIEW

Rodrigo Capobianco Guido and his friends in 2005, proposed a technique in which they used lossy compression algorithm that considered perceptual and rate distortion criteria. In their novel technique they achieved perceptually transparent coding of high quality audio sampled at 44100 Hz on 16 bits PCM, resulting 54 to 64 kbps bit rates. The decoder implements this algorithm effectively in real time [6].
Speech can be compressed with respect to its dynamic range and/or spectrum. The dynamic range used in telephones with A-law or u-law is capable of reducing 12 bits to 8 bits only; i.e. 96 kbps reduced to 64kbps. Speech compression can be achieved by scalar or vector quantization. This coding can be achieved using: (i) parametric or source method (CELP) (ii) Waveform method (ADPCM) produce compression rates above 32 kbps (iii) transform method coding (wavelet and fractal coding) produce compression rates 2.4kbps or even lower. Code Excited Linear Predictive (CELP) is one of the better compression schemes belonging to parametric scheme, can compress speech from 64kbps to 4.8 kbps [7].
Audio in its digital form provides much flexibility in audio functionalities. Digital audio compression allows the efficient storage and transmission of data and having 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 PCM. The digitization begins by sampling the audio input and quantizing the each sampled values into a discrete number with sampling rates from 8 kHz to 48 kHz. 8 kHz can cover most of the frequencies of human voice on 8 to 16 bits. He concluded that these approaches apply, low complexity, low compression and medium quality algorithms to audio signals [8].
Bartek Plichta and Mark Kornbluh, made research on "Digitizing Speech Recordings for Archival Purposes" and described digitizing speech recordings for archival purpose. Analyzing frequency response, dynamic range, noise, psychoacoustic and perceptual quality, spectral evaluation of recordings was used to develop digitization best practices. Digitization with sampling rate: 96KHz; on 24-bits; The WAV file format is recommended for speech recordings and are easy to process into a variety of streaming formats [9].

III. MATERIALS AND METHODS

The novel approach is proposed on using PCM coded speech, sampled at 8 KHz and quantized with 8 bits per sample. The coded speech is then transmitted from encoding to decoding end by dropping 3 least significant bits of each and every byte. A codebook containing 8 bit patterns for these dropped 3 least significant bits is used to complete the byte at the decoding end. From the codebook, bit patterns are randomly picked one by one and substituted as the least significant bits of each and every byte. The algorithm achieves 37.5% compression with negligible loss of speech quality [10].

The approach also checked the Affects upon two most popular Pakistani languages entitled as "Urdu and Sindhi Language Perceptions Using Codebook Based Digital Speech Compression" [11].

Reducing the codebook sizes containing bit patterns for 3 least significant bits, refines the cascaded approach. This strategy achieves good speech quality i.e., for 3 bit substitutions MOS achieved is 4.22, compatible to the recent standard coding schemes based on complex algorithms. All of these codebooks of reduced sizes are used one by one for the random substitution of 3 bits dropped at encoding end in different experimentations, as shown in Fig.1.

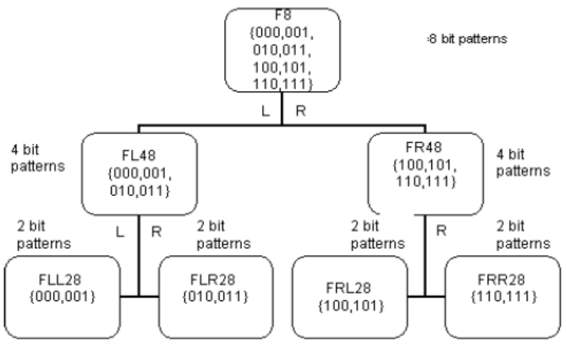

Fig.1. Reduced codebooks splitting 8 bit patterns for 3-bit substitution

The given Figure-1 is showing Filenames of speech files after substitutions of bit pattern and the original and reduced codebooks (containing bit patterns) in each box.

For the cascaded approach after random substitutions of bits from the reduced codebooks Error Analysis of bytes is conducted. Analyzing bit errors, if occurs in a byte are determined in Table-2.

Table-2 Bit Error in a byte substituting 3 least significant bits

0-bit error	1-bit error	2-bit error	3-bit error
0.0	12.5	25.0	37.5

- If the bits are same before and after substitution, then it is 0%-bit error in that byte.
- 1 different bit out of 8 and 7 bits are same in a byte then that error is 1-bit error and it is 12.5% error of that byte.
- 2 different bits out of 8 and 6 bits are same in a byte then that error is 2-bit error and it is 25% error of that byte.
- 3 bits out of 8 are different in a byte then that error is 3-bit error and it is 37.5% of that byte.

IV. RESULTS AND DISCUSSIONS

After random substitution of 3 least significant bits from the original/reduced codebooks, speech files showing bit errors in Table-3 (a).

Table- 3 (a) Error Free Bytes and Average Error in Bytes after 3 Bit Substitution

File Name	Bit Patterns	1-bit error	2-bit error	3-bit error	Error Free Bytes	Avg Error	% Error Occurring Bytes
F8	{0,1,2,3,4, ,6,7};	37.73	37.33	12.46	12.48	25.00	87.52
FL48	{0,1,2,3};	49.97	25.00	0.00	25.04	18.75	74.96
FR48	{4, 5,6,7};	24.91	50.12	24.97	0.00	25.00	100.00
FLL28	{0,1};	49.90	0.00	0.00	50.00	12.50	49.90
FLR28	{2,3};	49.82	50.18	0.00	0.00	18.75	100.00
FRL28	{4, 5};	50.83	49.97	0.00	0.00	18.75	100.00
FRR28	{6,7};	0,00	50,00	50.00	0.00	31.25	100.00

From all its 8 bit patterns at receiving side, 12.48% error free bytes i.e. bytes remained similar before and after substitution of bit patterns from the fixed codebook are obtained. While the rest of 87.52% bytes changed after substitution of bit patterns and average error recorded in every byte of the whole speech file named F8, was 25%, which means F8 is having 2-bit error in almost all of its bytes.

From the 2 Fixed reduced codebooks containing 4 bit patterns each at receiving side, codebook {0,1,2,3} gives 25.04% error free bytes 74.96 bytes having 18.75% error, it means that speech file named FL48 (File created at left side of the original having 4 bit patterns out of 8 bit patterns as shown in Fig-1) containing 1-bit to 2-bit error occurring bytes, codebook {4,5,6,7} gives 0% error free bytes and 100% error occurring bytes having 25% error in each of them, it means that speech file named FR48, containing 2-bit error occurring bytes.

From the 4 Fixed reduced codebooks containing 2 bit patterns each at receiving side, Codebook {0,1} gives 50.10% error free bytes 49.90 bytes having 12.5% error, it means that speech file named FLL28, containing 1-bit error bytes, Codebook {2,3} bit patterns gives 0% error free bytes and 100% error occurring bytes having 18.75% error in each of them, it means that speech file named FLR28, containing 1-bit and 2-bit error bytes. Codebook {4,5} bit patterns gives 0% error free bytes and 100% error occurring bytes having 18.75% error in each of them, it means that speech file named FRL28, containing 1-bit and 2-bit error bytes. Codebook {6,7} gives 0% error free bytes and 100% error occurring bytes having 31.25% error in each of them, it means that speech file named FRR28, containing 2-bit and 3-bit error bytes.

IV. CONCLUSION

The algorithm works on reduced codebooks for the random substitution of bit patterns. It is observed that shorter the codebook size i.e., number of bit patterns and lower the numeric values it contains, the better the result is. The larger the codebook size and higher the numeric values it contains, the worst the result will go. For instance if 2 bit pattern codebook with the numeric values 0,1 are taken for random substitutions of bits, best results are achieved. Error free result is indirectly proportional to the codebook size and the numerical values it contains.

In other words the better the error free result is the shorter the codebook size and the lower values it contains. The worst the error free result is the larger the codebook size and the higher the values it contains. The data rates achieved are 40 Kbps for 3 bits substitutions with respective codebook sizes. The approach can be used for the 4 and 5 bits substitutions.

REFERENCES

[1] http://en.wikipedia.org/wiki/Pulse-code_modulation, 19 August 2007.

[2] http://www.cisco.com/warp/public/788/signalling/waveform_coding.html, Feb 02, 2006

[3] http://en.wikipedia.org/wiki/CELP, 30 June 2007.

[4] http://www.otolith.com/otolith/olt/lpc.html, October 95

[5] Ray Horak, “Telecommunications and data communications handbook”, Wiley-Interscience, 2007, ISBN 0470041412, 9780470041413, pp. 512.

[6] Rodrigo Capobianco Guido, Lucimar Sasso Vieira, Fabrycio Lopes Sanchez, Jan Frans Willem Slaets, Lyrio Onofre Almeida, Adilson Gonzaga, Marcelo Bianchi, A Matched FIR (Finite Impulse Response) Filter Bank for Audio Coding, pp. 796-801, Proceedings of the Seventh IEEE International Symposium on Multimedia (ISM’05)

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

[8] Devis Yen Pan, “Digital Audio Compression”, Digital Technical Journal vol. 5 No. 2, Spring 1993.

[9] Bartek Plichta, Mark Kornbluh, Digitizing Speech Recordings for Archival Purposes, Matrix, Michigan State University, 2001

[10] Nazish nawaz Hussaini and Prof. Dr. Asadullah Shah, “Binary codebook Based Digital Speech Compression”, SZABIST Journal of Independent Studies and Research, Vol. 5, No. 2, July 2007

[11] Nazish nawaz Hussaini and Prof. Dr. Asadullah Shah, “Affects upon Urdu and Sindhi Language Perceptions Using Codebook Based Digital Speech Compression”, proceedings of Shaikh Ayaz Conference on Language and Literature, 2008