Voice encoding for wireless communication based on LPC, RPE, and CELP

Mohamed Muthanna Al-Heeti
Department of engineering
University Of Anbar
Ramadi, Iraq
mohamed.muthanna@uoanbar.edu.iq

Jamal A. Hammad
Department of Clinical Laboratory Science
College of Pharmacy, University Of Anbar
Ramadi, Iraq
jamal.ali@uoanbar.edu.iq

line 1: 3rd Ahmed Shamil Mustafa
Department of Computer Engineering Techniques
*Al-Maarif University College*Rramadi, Iraq
ahmedshamil90@gmail.com

***Abstract*— Nowadays, voice encoders are one of the basic elements in the multimedia and telecommunications. This report explains properties of speech, how sound waves are generated and how they are classified and what is voice and unvoiced speech. A basic waveform encoding technique, Pulse Code Modulation (PCM) is introduced as the most common technique by changing an analogue signal to digital data. Three types of voice Encoder: Linear Predictive Coder (LPC), Regular Pulse – Excited (RPE) and Code-Excited Linear Predictive (CELP) are briefly discussed with Performance Comparison table between them is made to give differences summary. Finally, The Dynamic Rhyme Test (DRT) and Mean Opinion Score (MOS) methods are discussed as a voice quality tests to evaluate the performance of voice coders and checking the quality of the artificial speech.**

Keywords—Voice Encoder, Linear Predictive Coder (LPC), Regular Pulse Excited (RPE), Code-excited Linear Predictive (CELP).

I. INTRODUCTION

Nowadays, one of the most important factors in wireless communication is the voice. The objective of voice coding is to compute and pressure speech signals to low bit rate transmitted, with protecting the important features quality which are necessary for a specific application. As the physical spectrum of wireless communication is limited and service provider must expand a large number of users with a limited Bandwidth. The transmitted voice signal has to be changed to a form that ensures the Bandwidth in a way that will not affect the quality of original signal and voice encoding is that technique. Some other restrictions should be taken into consideration in coding signal such as, time delay, complexity and acting with packet losses or bit errors [1].

Speech is generated when air pressure, which comes from the lungs, forces the vocal cords to open and oscillate. Types of speech sounds based on the generation of a constriction at the lips or at the several point in the vocal tract. Figure (1) shows The Human Speech Production System. It is important here to explain what is meant by properties of speech .The target in speech synthesis is to switch a string of text, or a sequence of words, into natural sounding speech. Voiced and unvoiced are two types of speech sounds [2] [3].

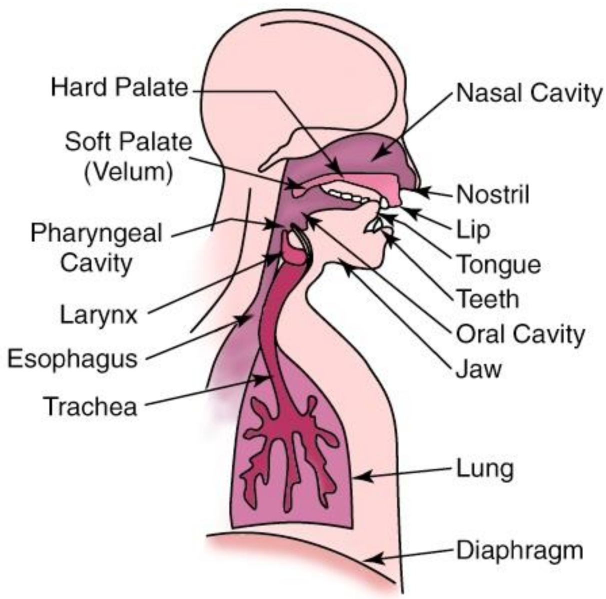

Fig. 1. The Human Speech Production System [2].

A. Voiced Speech

Voiced speech: as mentioned above, Speech is generated when air pressure, which comes from the lungs, forces the vocal cords to open and oscillate. The vibrational frequencies (pitch) change from (50 to 400) Hz (that depending on several thing such as the person's sex and age) and produces resonance in the voiced path at single harmonics. The peaks of these resonances are named formants [4] [3].

B. Unvoiced Speech

Unvoiced sounds are generated when forcing air over an opening, named fricatives are produced. Fricatives do not create many periodicities as in voiced because it does not oscillate the voiced Cords speech [2].

II. GENERAL ENCODING OF WAVEFORMS

It is very important to encode the waveform by converting it from analogue signal into digital form. In order to provide good quality, Waveform codecs are typically used at high bit rate. There are several examples of these general encoders contain Pulse Code Modulation for telecommunications and Uniform Binary Coding for music Compact Disks [5].

The most common technique using to change an analogue signal to digital data is called pulse code modulation (PCM). It is one of the waveform encoder technique that is used in popular voice rating circuits. Firstly, the signal bandwidth is determined by a low pass or band pass filter to get a specific band signal. switch or Sampler which works on a high rate of speed, is a first step which samples the signal into a sequence of pulses in the period of time between each pair of them this formula is called Nyquist theorem. The next step is quantization to produces error which can be calculated as a signal to quantization error. The actual error is the difference between the original sample amplitude and the quantized sample, which can be decreased by expanding the range of quantization levels. The final step is encoding, in which the number of bits are distributed to every quantization levels based on number of bit per sample [7] [10].

The PCM encodes eight bits per sample at 8 KHz sample rate. So the minimum bandwidth wanted is 64kbps, and the bit rate value is also equal to 64kbps, thus, the process of saving bandwidth is not considered as an efficient. Also, this process is not feasible throughout communication channels in case that the bandwidth characteristic is premium [3] [6].

III. TYPES OF VOICE ENCODERS

Digital voice communication systems use Vocoders to minimize transmitted bit rate by digitizing and compressing speech signals as shown in Figure (2). There are several modern types of speech-coding schemes use in mobile systems, the following are discussed LPC, REP and CELP operations [1].

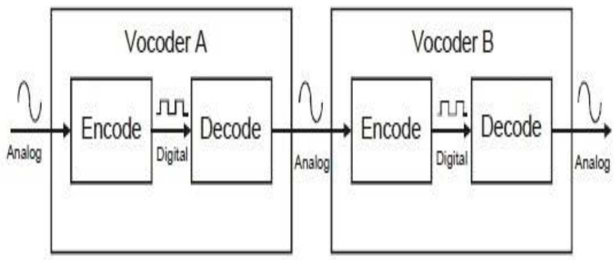

Fig. 2. Vocoder in a tandem configuration [1].

A. Linear Predictive Coders (LPC)

It is the most popular technique which is used in speech coding at low bit-rate. Lately, LPC has been considered as a very important tool especially when dealing with speech analysis. In analysing the speech, there are two extremely separate components the first one is the glottal excitation (LP excitation) and the second one is the voiced tract parameters (LPC coefficients) and both of these are resolved by LPC. [13] [8].

The operation of the LPC encoder splits into several steps as shown in figure (3); first a rectifier and 20 Hz low pass filter using to sample the input speech. After that the block of samples which received from audio waveform (segment) is analysed. Then, determining the perceptual characteristic of the waveform. Next the sampled segment is quantized and computing error signal .finally, the new set of coefficients is sent for each quantized section in a string shape of frames. LPC decoder receives the voiced, unvoiced, amplitude and pitches signals from LPC encoder. By giving it to the balance modulator it processes. These are modulated by the carrier, created either by noise generator or pulse generator depending upon the unvoiced or voiced sounds respectively. The output from that is input to the adjustable filter. The filter parameter adjusting signals are also received and are used as at the encoder to get optimum voice regeneration. The model of LPC Decoder shown in Figure (4) has several parameters. These parameters are voicing, Gain, Filter coefficients and Pitch period [14] [19].

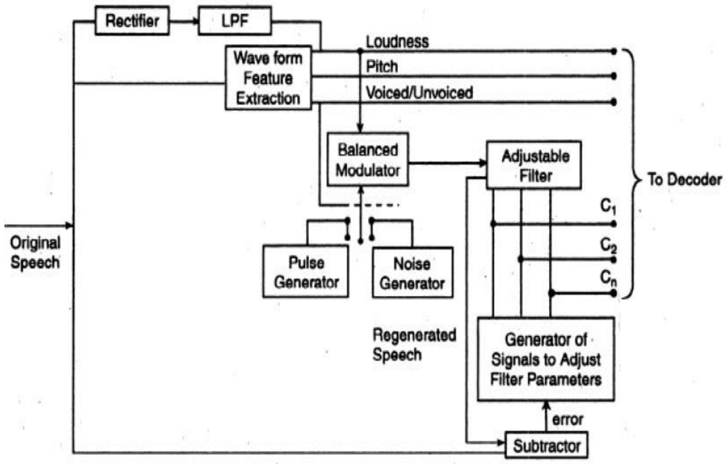

Fig. 3. LPC Encoder [19].

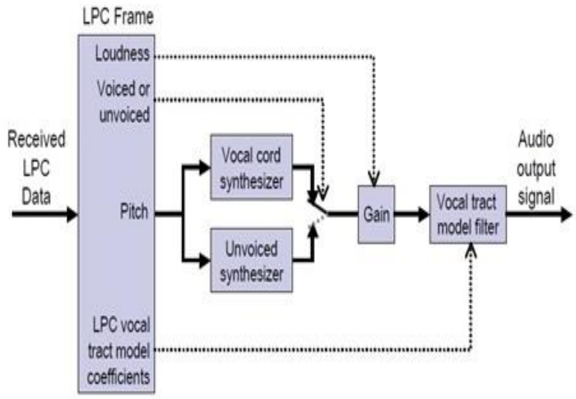

Fig. 4. The model of LPC Decoder [11].

The main disadvantage of the LPC is the speech quality is very low, but this can be improved by using hybrid coders [12].

B. Regular pulse-Excited (REP)

There is another type of hybrid coding is Full rate Linear Predictive Coding - Regular Pulse Excited (LPC-RPE). This type is used to produce balance between data rate and the quality of GSM network. As a traditional PCM the similar sample rate 8000 sample/sec is used to represent the input source which contains thirteen bits for each sample. There are several important steps are needed to generate 13 kbps. Firstly, the input samples are divided into block and each block contents into 160 samples. After that it is applied to a short term of LPC, where 36 bits per block is output of the filter coefficients of LPC. These blocks are splitted into 4 sub stages of 40 samples and provided that to the LTP (Long Term Prediction) filter to get the coefficient to represent the pitch gain and period. The overall is 36 bits for 160 samples due to each sub block represents by 9 bits. [20].

RPE algorithm used to compress The output of LTP in which each Block explain by one of the 4 applicants that means it contains 13 samples for each one. Grid position of RPE is the standard of selection. Finally, the overall operation gives 13 Kbps. [3].

C. Code-Excited Linear Prediction (CELP)

There are many attempt to improve the existent LPC coder and the first idea of code-excited linear prediction (CELP) was created by Schroeder and Atal in 1985.It is one of the most effective Coding methods at low bit-rates. The gap between vocoders and waveform coders could be bridged by producing low rate coded speech which is done by CELP algorithm similar to that of medium rate waveform coders [15].

Speech is splitted into several frames usually the length of each one between 10 to 30 ms. also an optimum set of linear prediction for each frame and determining and quantizing for the pitch filter parameters, figure (5) shows CELP coders .in CELP is an efficient closed loop analysis synthesis method for medium 16 kbps and low 4 kbps band speech coding systems. The CELP coders have a good reproduction of the speech signal because every frame is furthermore, splitted into several of subframes the length of each one is five ms and for every subframe an excitation codebook is checked to get the input vector to the quantized predictor system [16].

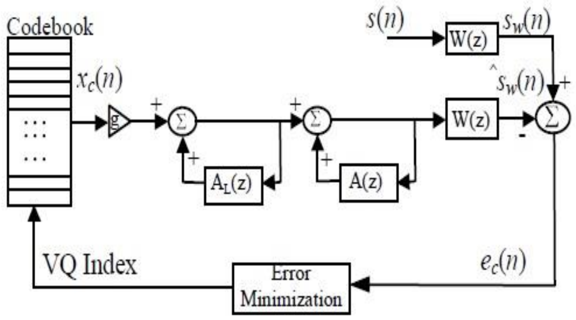

Fig. 5. Analysis-by-Synthesis CELP Encoder [12].

Short-term LP analysis was performed on each frame to make the LPC. Then, long-term analysis will used to each subframe. Usually the original speech, or pre-emphasized speech is the input to short-term LP analysis while the prediction error is the input to long-term LP analysis. After that altered format synthesis filter are processed by the pitch synthesis filter. The length of each excitation code vector and that of the sub-frame should be have the same length that means an excitation codebook search is worked on each subframe. For each sequence the mean-square error studied then the lowest error find in the code vector and gain .Finally, before the signal of excitation codebook, LPC, long-term LP parameters and gain will be transmitted as the CELP bit- stream it will be encrypted and packed [17].

The main disadvantage it is their significant computational requirements, but it still shows a perfect performance at low data rates [18].

D. Performance comparison

The table below shows main differences between LPC, RRE and CELP coders.

TABLE I. PERFORMANCE COMPARISON [28]

	LPC	**RPE**	**CELP**
Data rate	2.4 Kbps	13 Kbps	4.8 Kbps
Distinction between Voice & Unvoiced	Uses switch	Uses regular spaced pulse	Identifies voiced and unvoiced Speech
Speech Quality	Poor	Relatively good	Very good

IV. VOICE QUALITY MEASUREMENTS

Cole proposed that the voice quality, communication delay, data rate and coding algorithm complexity are the most important quality measures. The conveniently scale voice quality is based on several things such as test data, listening groups and the speaker. Thus, information from one test group must not be contrasted with others. There are many commonly tests used such as Dynamic Rhyme Test (DRT) and Mean Opinion Score (MOS) [21] [22].

A. The Diagnostic Rhyme Test (DRT)

In 1958, Fairbanks presented The Diagnostic Rhyme Test (DRT). It was designed to test the clearness of coders known to present speech of lower quality due to the listener must define which consonant was spoken when given with a pair of rhyming words, then Rhyme tests are so named. It consists of 96 word pairs and these are selected to measure the pronunciational characteristics. The listener has required to recognize among word pairs such as dense – tense, caught-taught, show - sow, and pool-tool. Finally, DRT is a test method used in a widely form and it is very useful as a developing tool, but the main weakness it does not test all possible interferences among consonants [25].

B. Mean Opinion Score (MOS)

The MOS test is one of the most popular private tests validate the speech quality processing systems , so it is the most widely used ,MOS test describes speech intelligibility and it is used in voice telecommunications , furthermore it is proper for overall evaluation of synthetic speech [25][26] .

Mean opinion score (MOS) is one of a type of test that is used to obtain the quality of a voice sample according to listeners judge. Absolute Category Rating (ACR) as the table below is type of the MOS. this type, there are five categories regarding the quality and the hearers are asked to choose a single option for each tracks of the voice and the average value give the MOS value [29].

TABLE II. MEAN OPINION SCORE [29].

Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

V. CONCLUSION

To sum up, it can be clearly seen that transmitting digital voice has many important benefits; much bandwidth can be saved by pressing the signal and decreasing the bit rate .PCM was the first encoder design to achieve like a transformation. The performances of different encoders are commonly compared by considering the voice quality, data rate and intelligibility. DRT and MOS techniques are used to measure the quality of coders. MOS is the most frequently method used. Several modern types of voice encoders (LPC, REP and CELP) are explained, although of their low data rate, operators still depend on them.

REFERENCES

[1] NTIA Report 01-386 - Voice Quality Assessment of Vocoders in Tandem Configuration, Christopher Redding, Nicholas DE Minco, and Jeanne Lindner, U.S. DEPARTMENT OF COMMERCE, April 2001.

[2] Purdue University: ECE438 - Digital Signal Processing with Applications - Speech Processing - Labs - Proof Charles A. Bouman.

[3] B. Douglas, 'Voice Encoding Methods for Digital Wireless Communications Systems', 2014. [Online]. Available: http://lyle.smu.edu/~levine/ee8320/BDouglas.PDF. [Accessed: 09- Oct- 2014].

[4] McElroy, Ciarán. “Speech Coding”. http://wwwdsp.ucd.ie/speech/tutorial/speech_coding/speech_tut.html (October 1997).

[5] A. Das, Digital communication. Berlin: Springer, 2010.

[6] LucentTechnologies, Technology Description, Http://standards.lucentssg.com/description.html.

[7] The T1 Carrier System', Bell Labs Rec, vol. 40, no. 10, pp. 358--364, 1962.

[8] Speech Synthesis and Recognition by J. N. Holmes and Wendy Holmes 2nd edition.

[9] Lecturer Mark Handley, Chapter4, Speech Compression, http://www.cs.ucl.ac.uk/teaching/ Z24/.

[10] B. Forouzan and S. Fagan, Data communications and networking. New York: McGraw-Hill Higher Education, 2007.

[11] Lund, Anders, and Theodore S. Rappaport, 'Wireless Communications – Principles and Practice Second Edition'. (2012): Web. 10 Dec. 2015.

[12] I. Atkinson, S. Yeldener, and A. Kondoz, “High quality split-band LPC vocoder operating at low bit rates,” in Proc. Int. Conf. on Acoust. Speech, Signal Processing, May 1997.

[13] Digital Voice INC, 'Coding Overview', 2011. [Online]. Available: http://www.dvsinc.com/papers/vc_over.htm. [Accessed: 10- Oct- 2014].

[14] Fundamental of Voice –Quality Engineering in Wireless Networks by Avi Perry.

[15] Chu, Wai C. Speech Coding Algorithms. Hoboken, N.J.: J. Wiley, 2003. Print.

[16] Li, Wei, Anu Sridhar, and Tina Teng. 'Comparison of Speech Coding Algorithms: ADPCM, CELP and VSELP'. (1999).

[17] Digital Speech, By Ahmed M. Kondoz, Second Edition.

[18] A. Gerson, Ira, and Mark A. Jasiuk. 'VECTOR SUM EXCITED LINEAR PREDICTION (VSELP) SPEECH CODING AT 8 KBPS'. IEEE (1996).

[19] R. Avudaiammal, Information coding techniques. New Delhi: Tata McGraw Hill Education, 2010.

[20] N. Bhatt, Y. Kosta and V. Tank, 'Proposed Modification in ETSI GSM 06.10 Full Rate Speech Codec for High Rate Data Hiding and its Objective Evaluation of Performance Using Simulink', in Communication Systems and Network Technologies (CSNT), 2011 International Conference on, Katra, Jammu, 2011, pp. 27 - 33.

[21] Cole, Ronald A, et.al. Editors. Survey of the State of the Art in Human Language Technology. http://www.cse.ogi.edu/CSLU/HLTsurvey (November 1995, last visited November 1997). Reference from chapter 10.1, written by Isabel Trancoso, and 10.2, written by Bishnu S. Atal and Nikil S. Jayant.

[22] Dynastat Corporation. "Summary of Speech Intelligibility Testing Methods". http://www.realtime.net/dynastat/intellig.html (last visited January 1998).

[23] D. Gibson, Jerry. 'Speech Coding Methods, Standards, and Applications'. Santa Barbara, CA 93106-6065.Department of Electrical & Computer Engineering University of California, Santa Barbara.

[24] The Electrical Engineering Handbook by Richard C. Dorf, 2nd edition.

[25] http://www.ukessays.com/essays/communications/the-basic-properties-of-speech-communications-essay.php#ixzz3snYus0YR .

[26] Speech Analysis & Synthesis Methods developed at ECT in NTT by N. Sugamura and F. Itakurna, 1986.

[27] Larijani, Hadi, and Kapilan Radhakrishnan. 'Voice Quality in Voip Networks Based On Random Neural Networks'. Glasgow Caledonian University G4 0BA (2010).

[28] Cl.cam.ac.uk, 'The ITU (was CCITT) Audio Family', 2014. [Online]. Available: http://www.cl.cam.ac.uk/~jac22/books/mm/book/node109.html. [Accessed: 15- Nov- 2014].

[29] Cisco, 'Understanding Codecs: Complexity, Hardware Support, MOS, and Negotiation', 2014. [Online]. Available: http://www.cisco.com/c/en/us/support/docs/voice/h323/14069-codec-complexity.html#mos. [Accessed: 15- Nov- 2014].