Investigation of Speech Coding Algorithms

Andon Nenov[1], Georgi Iliev[2], Maria Nenova[3]

Abstract – **In this paper, an investigation of different types of speech coding algorithms is presented. The goal of speech coding systems is to transmit speech with good quality. Different types and specifics of algorithms and their applications are shown. Features of Adaptive Differential Pulse Code Modulation, Code Excited Linear Predictive and MELP coders are presented too.**

The complexity of the algorithm is also a very important point for research. Since speech compression is used in real-time systems, digital signal processors are the best choice for running the algorithm. Implementations on DSP-based systems are not only robust and flexible but also very powerful.

Keywords – **Speech coding algorithms, Coders.**

I. Introduction

The speech coding is generally used for representation of the human voice as a digital signal. The main advantage of use of coding the speech signals is the ability to compress the signal, in a sense to reduce the bit rate of the digital speech signals [1].

The goal of all speech coding systems is to transmit speech with the highest possible quality using the least possible channel capacity. It is important to investigate the bit-rate and the relation with algorithmic complexity for the speech coders. Speech compression is used in real time systems and thus they are implemented on DSP-based systems. The algorithm complexity must be low, due to the power consumption requirements.

During the years development of coders operating at 4,8 kbits/s and below for narrow band and secure transmission in communications and voice applications in Internet arises [2]. Unfortunately the characteristics of those type low data-rate coders are poor. [3][4]

Other type of algorithms is MELP [2]. The MELP vocoder described in [5] is an enhancement of CELP with a number of additional features. These features include a mixed excitation signal (i.e., a mixture of noise and pulse excitation used as input to the synthesis filter), an adaptive spectral enhancement filter, Fourier magnitude modelling of the pulse excitation, and a pulse dispersion filter. The MELP vocoder encodes 22.5 ms of speech to a 54-bit frame.

[1]Andon Nenov is with the Ministry of Inferior, E-mail: and_mail@mail.bg

[2]Georgi Iliev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gli@tu-sofia.bg

[3]Maria Nenova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: mvn@tu-sofia.bg

II. Coder Implementation

A. Problem Statement

In mobile communication systems one of the main problem is the limited bandwith. Therefore speech coders providing the quality of speech signals at low bit rate are needed. The main objective of the paper is to compare some of the most commonly used algorithms in wireless communication systems: CELP, VSELP, MELP and ADPCM.

Most of the already developed in this manner coders have been adopted in cellular phone standards. The focus is on those coders because of the fact that in UMTS and CDMA systems they are implemented.

On Fig. 1 is depicted one of the possibilities of classification of speech coders:

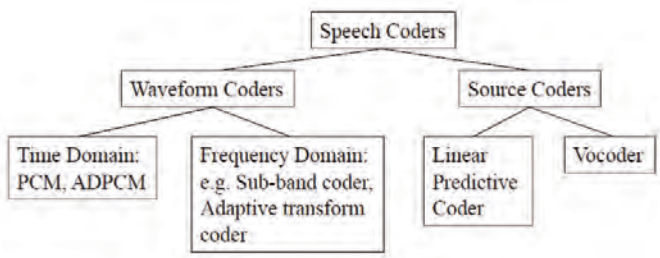

Fig. 1 Varieties of speech coders

A comparison of those types of coders, called by some researcher vocoders is going to be made in the next sections. The structure of coder and decoder and its components will be investigated.

B. ADPCM coders

The coders group called waveform coders can be investigated in the time domain and in the frequency domain. Specific their feature is that they are trying to achieve the time waveform of the speech signals. One of the biggest their advantages is that they are robust for the most of the speech characteristics and are very useful for applications in noisy environment.

The waveform codecs algorithm of work is based on the ability without knowledge of the prior information about the coded signal, to produce a reconstructed signal whose waveform is close to the form of the original signal. Their biggest advantage is that the low complexity. When the data rate is lower than 16 kbits/s the reconstructed speech quality that can be obtained degrades rapidly.

One of the main classes in the time domain group of algorithms is Adaptive Differential Pulse Code Modulation (ADPCM). ADPCM codecs are waveform codecs which

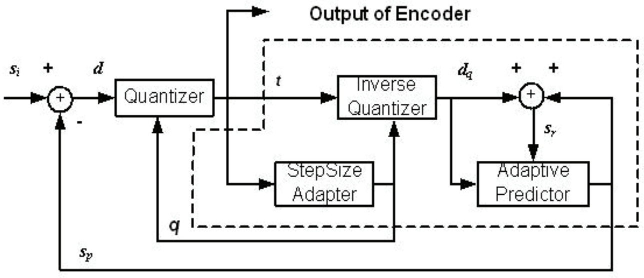

Fig. 2: ADPCM Coder Block diagram

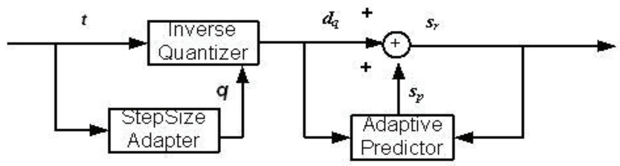

Fig. 3: ADPCM Encoder Block diagram

instead of quantizing the speech signal directly, like PCM codecs, quantize the difference between the speech signal and the prediction that has been made of the speech signal. If the prediction is accurate then the difference between the real and predicted speech samples will have a lower variance than the real speech samples, and will be accurately quantized with fewer bits than would be needed to quantize the original speech samples. At the decoder the quantized difference signal is added to the predicted signal to give the reconstructed speech signal. The performance of the codec is aided by using adaptive prediction and quantization, so that the predictor and difference quantizer adapt to the changing characteristics of the speech being coded.

The CCITT standardized a 32 kbits/s ADPCM, in G.721, which gave reconstructed the speech signal as in the 64 kbits/s PCM codecs. Later in recommendations G.726 and G.727 codecs operating at 40,32,24 and 16 kbits/s were standardized.

The ADPCM algorithm for compression of the signal is shown can be implemented with the blocks depicted on Fig.2. The process is on the base of iterations. On the next iteration, the predicted sample s_p and the quantizer step size index are saved in a structure.

The quantization step size and the predicted sample s_p are initially set to zero. The input s_i to the speech encoder is supposed to be a 16-bit 2's complement speech sample, while the value returned by the speech encoder is an 8-bit number which contains the 4-bit sign magnitude ADPCM code.

On the Fig. 3 the signal of the difference d is produced by subtracting the predicted sample s_p from the input signal s_i. Then the signal d is fed to the quantizer and adaptive quantization is performed on the difference obtained in the previous step.

One of the advantages of the structure on Fig. 2 is that within the encoder there is a decoder inside it. This ensures synchronization between encoder and decoder without requiring any additional data. The dotted lines shown in Fig.2 show the block comprising the embedded decoder.

The ADPCM value is used by the embedded decoder to update the inverse quantiser, which in turn produces a dequantized version d_q of the difference d. To simplify the speech compression process a fixed predictor has been used instead of an adaptive predictor, which significantly reduces the amount of data memory and instruction cycles required. A weighted average of the last six dequantized difference values and the last two predicted values are used by the adaptive predictor of ITU G.721 for its adjustment and updating according to the value of each input sample. At this point, new predicted sample s_r is obtained by adding the dequantized difference d_q to the predicted sample s_p. Finally, the new predicted sample s_r is saved in s_p.

D. CELP Codec **(***Hybrid codecs)*

Usually compression methods are based on entropy coding or on source coding. The entropy coding is also called lossless coding. If entropy coding is implemented then there is used redundancy in order to decrease the amount of the data to be compressed.

In case of information filtered out to reduce the unnecessary elements, the coding is called source coding and this I s the most commonly used for compression of video and audio streams. A typical feature of source coding is the loss of information, which means that the decompressed data stream will not contain all the elements of the original information. One of the best examples of coding with losses is MPEG coding [3].

The so called Hybrid codecs attempt to operate between waveform and so called source codecs. Waveform codecs are capable of providing good quality speech at bit rates about16 kbits/s, but are of limited use at rates below this. Source codecs on the other hand can provide understandable speech at 2.4kbits/s and below, but cannot provide natural sounding speech at any bit rate.

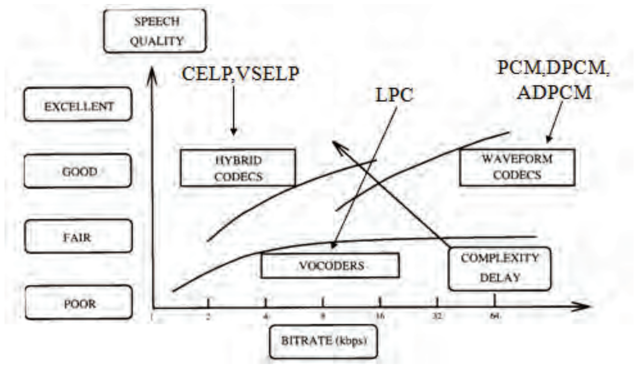

Fig. 4: Speech versus bit-rate classification of speech codecs

As can be seen from the figure 4, hybrid codecs combine techniques from both source and waveform codecs and as a result give good quality with intermediate bit rates.

The coding of this algorithm is based on analysis-by-synthesis search procedures, vector quantization (VQ) and linear prediction (LP).

The one of the most commonly used hybrid codec type is Analysis-by-Synthesis (AS) codecs. Such coders use the same linear prediction filter model as found in source codecs. However instead of applying a simple two-state, voiced/unvoiced, model to find the necessary input to this filter, the excitation signal is chosen by attempting to match the reconstructed speech waveform as closely as possible to the original speech waveform. Thus AbS codecs combine the techniques of waveform and source codecs.

The principle of work of AS codecs is splitting the input speech to be coded into frames. For each frame parameters are determined for a filter called synthesis filter. The excitation to this synthesis filter is determined by finding the excitation signal, which minimizes the error between the input speech and the reconstructed speech. Thus the name Analysis by Synthesis means that the encoder analyses the input speech by synthesizing many different approximations to it. The basic idea is that each speech sample can be approximated by a linear combination or the preceding samples.

One more type of low bit rate coders implemented in communications is called CELP (Code Excited Linear Predictive). The CELP was first introduced by Atl and Schroder [5]. In order to achieve real-time encoding, the CELP optimisation is divided into smaller, sequential searches using the perceptual weighting function described earlier[6][7][8]. CELP is based on vector quantization. One of the most commonly used algorithm for producing good quality speech at rates below 10kbits/s is CELP.

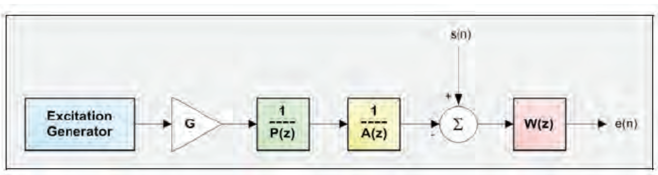

Fig. 5 Scheme of CELP encoders

On the Fig.5 is depicted a block diagram of a CELP encoder. The generator produces sequence which will be multiplied and then given to a filter called pitch synthesizing filter *1/P(z)*.

The received signal is then fed to one other filter with transfer function *1/A(z)*, where

$$H(z) = \frac{1}{A(z)} = \frac{G}{1 + a_1 z^{-1} + a_2 z^{-2}},$$

$$A(z) = 1 + \sum_{k=1}^{p} a_k . z^{-k}$$

The a_k in the formula is coefficient called the Linear Predictive Coefficient (LPC). The coefficients are determined by minimising difference between actual signal and predicted signal by the use of least square method. The variable p gives the order of the filter. This filter is intended to model the short-term correlations introduced into the speech by the action of the vocal tract. This kind of coding is also called Linear Predictive Coding (LPC)

The difference between the speech signal and the original speech spectrum *S(n)* is then weighted according to a subjective error criterion, *W(z)* to receive the error signal *e(n)*, which is encoded using vector quantization.

Often CELP is called a hybrid codec because it uses both waveform and source coding techniques. Unfortunately during the process of filter coefficients updates is introduced a high rate of delays. For a typical hybrid codec this delay will be of the order of 50 to 100 ms, and such a delay lead to problems. Thus, many efforts have been focused in providing a standard codec that has as bit rate 16 Kbps, while providing a quality comparable to that provided by the ADPCM 32 Kbps. The major challenge is to reduce the delay to about 5 ms. CELP needs a dedicated hardware to run in real time.

The disadvantage of CELP coding schemes is that they fail to represent the high frequencies in speech at bit rates around 6 kbit/sec or lower. For this reason, the newer CELP schemes are actually combinations of CELP and MLPC (Multipulse LPC), using a limited number of excitation pulses.

There are many papers on the latest CELP/MLPC coders which provide a deeper inside to their operation. The CELP needs a fixed bitrate of only 4.8 kbit/sec for encoding human speech, but it is worth mentioning that among all low-bitrate vocoders CELP demands the highest computational power.

The MELP vocoder described in [9] is an enhancement of CELP with a number of additional features. These features include a mixed excitation signal (i.e., a mixture of noise and pulse excitation used as input to the synthesis filter), an adaptive spectral enhancement filter, Fourier magnitude modelling of the pulse excitation, and a pulse dispersion filter. The MELP vocoder encodes 22.5 ms of speech to a 54-bit frame. Most of the parameters in a typical frame are quantized using suitable vector codebooks, and only the codebook indices are transmitted. MELP is much faster than CELP, but needs at least 4 times as much memory as CELP does.

The disadvantage of CELP coding schemes is that they fail to represent the high frequencies in speech at bit rates around 6 kbit/sec or lower. For this reason, the newer CELP schemes are actually combinations of CELP and MLPC (Multipulse LPC), using a limited number of excitation pulses.[9]

F. VSELP Codec (Hybrid codecs)

There is also one more important type codec with implementation in GSM communications called a half-rate GSM codec. It is a Vector Self-Excited Linear Predictor (VSELP) codec at bit rate of 5.6kbit/s. VSELP codec is a close relative of the CELP codec family explained in the previous chapter. A slight difference is that VSELP uses more than one separate excitation codebook, which is separately scaled by their respective excitation gain factors.

TABLE 1. PERFORMANCE AND COMPLEXITY OF ALGORITHMS

Algorithm	Bit Rate (bit/sec)	MIPS
PCM	64 k	0.01
ADPCM	32 k	2
CELP	4.8 k	16
VSELP	8 k	13.5

In Table 1 is presented comparison between different types of the already investigated algorithms with their bit rate. Implementing the original systems required several hundred MIPS (Million Instructions Per Second).

As the bit rate reduces the computational complexity increases. Much of the research being done has concentrated on reducing this load in order to facilitate implementing the algorithm on available silicon.

III. CONCLUSION

Many communication channels suffer from noise, interference or distortion due to hardware imperfections, or physical limitations. The goal of error control coding is to encode information in such a way that even if the channel (or storage medium) introduces errors, the receiver can correct the errors and recover the original transmitted information.

This report discusses algorithms such as ADPCM and CELP, MELP, VSELP and other encoding methods. It also discusses the technological advances. Speech can be coded at many levels. Lower bit-rates are achieved by imposing more constraints of the speech production mechanism are applied. The failure becomes more catastrophic as the bit rate is reduced.

REFERENCES

[1] Wai C. Chu, *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders*, John Wiley & Sons, Inc., 2003.

[2] Kohler M.A., “A Comparison of the New 2400BPS MELP Federal Standard with other Standard Coderds, IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP’97), Munich, Germany

[3] Spanias, A.S. (1994). Speech coding: a tutorial review. In: *Proceedings of the IEEE* Vol. 82, October 1994, pp. 1541-1582.

[4] Rabiner L.R. (1994). Applications of Voice Processing to Telecommunications. In: *Proceedings of the IEEE* Vol. 82 No. 2, February 1994, pp. 174-180.

[5] Schroeder M.R. and Atal B., “ Code Exited Linear Prediction (CELP): High Quality Speech at Very Low bit Rates”, Proc. ICASSP-85, P. 937, Tampa, Apr. 1985

[6] Liang, C., and Xiongwel, Z. (1994). A New 1.2kb/s Speech Coding Algorithm and its Real-time Implementation on TMS320LC548. *Institute of CommunicationsEngineering, PLA University of Science and Technology*, Nanjing

[7] Kohler M.A., “A Comparison of the New 2400BPS MELP Federal Standard with other Standard Coderds, IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP’97), Munich, Germany

[8] TEXAS INSTRUMENTS (2002). *TMS320C6000 Programmer’s Guide*. http://focus.ti.com/lit/ug/spru198g/spru198g.pdf.

[9] U.S. DEPARTMENT OF DEFENSE (1998). *Specifications for the Analog to Digital Conversion of Voice by 2,400 Bit/Second Mixed Excitation Linear Prediction*