G.729, G.723.1, G.722.2 ITU-T Speech Codecs: A Review

Harsh Patel[1] Prof.Kinnar Vaghela[2]
[1]Research Student [2]Professor
[1,2]Department of Electronics & Communication Engineering
[1,2]L.D College of Engineering & Tech, Ahmedabad, Gujarat, India

Abstract—*The International Telecommunication Union coordinates the shared global use of the radio spectrum, promotes international cooperation in assigning satellite orbits, works to improve telecommunication infrastructure in the developing world, and assists in the development and coordination of worldwide technical standards. The ITU is active in areas including broadband Internet, latest-generation wireless technologies, aeronautical and maritime navigation, radio astronomy, satellite-based meteorology, convergence in fixed-mobile phone, Internet access, data, voice, TV broadcasting, and next-generation networks. One of the major application of International Telecomminication Union is standardization and developing speech codec according to required application. The basic function of speech codec is a speech coding. Speech coding is an application of data compression of digital audio signals containing speech. Speech coding uses speech-specific parameter estimation using audio signal processing techniques to model the speech signal, combined with generic data compression algorithms to represent the resulting modeled parameters in a compact bitstream. The two most important applications of speech coding are mobile telephony and Voice over IP. The techniques employed in speech coding are similar to those used in audio data compression and audio coding where knowledge in psychoacoustics is used to transmit only data that is relevant to the human auditory system. This review paper include description of speech codec G.729, G.723.1, G.722.2.*

Keywords—G.729, G.723.1, G.722.2, CELP, Linear Prediction

I. Introduction

The G.729 coder has algorithm for the coding of speech signals at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP). This coder is designed to operate with a digital signal obtained by first performing telephone bandwidth filtering of the analogue input signal, then sampling it at 8000 Hz, followed by conversion to 16-bit linear PCM for the input to the encoder. The output of the decoder should be converted back to an analogue signal by similar means. Other input/output characteristics should be converted to 16-bit linearPCM before encoding, or from 16-bit linear PCM to the appropriate format after decoding[8].

The G.723.1 is used for compressing the speech or other audio signal component of multimedia services at a very low bit rate as part of the overall H.324 family of standards. This coder has two bit rates associated with it: 5.3 and 6.3 kbit/s. The higher bit rate has greater quality. The lower bit rate gives good quality and provides system designers with additional flexibility. Both rates are a mandatory part of the encoder and decoder. It is possible to switch between the two rates at any frame boundary. An option for variable rate operation using discontinuous transmission and noise fill during non-speech intervals is also possible. This coder was optimized to represent speech with a high quality at the above rates using a limited amount of complexity. It encodes speech or other audio signals in frames using linear predictive analysis-by-synthesis coding. The excitation signal for the high rate coder is Multipulse Maximum Likelihood Quantization (MP-MLQ) and for the low rate coder is Algebraic-Code-Excited Linear Prediction (ACELP). The frame size is 30 ms and there is an additional look ahead of 7.5 ms, resulting in a total algorithmic delay of 37.5 ms[9].

The G.722.2 is a high quality Adaptive Multi-Rate Wideband (AMR-WB) encoder and decoder that is primarily intended for 7 kHz bandwidth speech signals. AMR-WB operates at a multitude of bit rates ranging from 6.6 kbit/s to 23.85 kbit/s. The bit rate may be changed at any 20-ms frame boundary[10].

II. G.729

The G.729(CS-ACELP) coder is based on the code-excited linear prediction (CELP) coding model. The coder operates on speech frames of 10 ms corresponding to 80 samples at a sampling rate of 8000 samples per second. For every 10 ms frame, the speech signal is analysed to extract the parameters of the CELP model (linear prediction filter coefficients, adaptive and fixed-codebook indices and gains). These parameters are encoded and transmitted. The bit allocation of the coder parameters is shown in Table 1. At the decoder, these parameters are used to retrieve the excitation and synthesis filter parameters. The speech is reconstructed by filtering this excitation through the short-term synthesis filter, as is shown in Figure 1. The short-term synthesis filter is based on a 10th order linear prediction (LP) filter. The long-term, or pitch synthesis filter is implemented using the so called adaptive-codebook approach. After computing the reconstructed speech, it is further enhanced by a postfilter[8].

Parameter	Codeword	Subframe 1	Subframe 2	Total per frame
Line spectrum pairs	L0, L1, L2, L3			18
Adaptive-codebook delay	P1, P2	8	5	13
Pitch-delay parity	P0	1		1
Fixed-codebook index	C1, C2	13	13	26
Fixed-codebook sign	S1, S2	4	4	8

Codebook gains (stage 1)	GA1, GA2	3	3	6
Codebook gains (stage 2)	GB1, GB2	4	4	8
Total				80

Table 1: Bit allocation of the 8 kbit/s CS-ACELP algorithm (10 ms frame)

III. G.723.1

The G.723.1 coder has two bit rates associated with it. These are 5.3 and 6.3 kbit/s. The higher bit rate has greater quality. The lower bit rate gives good quality and provides system designers with additional flexibility. Both rates are a mandatory part of the encoder and decoder. It is possible to switch between the two rates at any 30 ms frame boundary. An option for variable rate operation using discontinuous transmission and noise fill during non-speech intervals is also possible. This coder was optimized to represent speech with a high quality at the above rates using a limited amount of complexity. Music and other audio signals are not represented as faithfully as speech, but can be compressed and decompressed using this coder. This coder is designed to operate with a digital signal obtained by first performing telephone bandwidth filtering (ITU-T Rec. G.712) of the analogue input, then sampling at 8000 Hz and then converting to 16-bit linear PCM for the input to the encoder. The output of the decoder should be converted back to analogue by similar means. Other input/output characteristics, such as those specified by ITU-T Rec. G.711 for 64 kbit/s PCM data, should be converted to 16-bit linear PCM before encoding or from 16-bit linear PCM to the appropriate format after decoding. The bitstream from the encoder to the decoder is defined within this Recommendation[9].

The coder is based on the principles of linear prediction analysis-by-synthesis coding and attempts to minimize a perceptually weighted error signal. The encoder operates on blocks (frames) of 240 samples each. That is equal to 30 ms at an 8 kHz sampling rate. Each block is first high pass filtered to remove the DC component and then divided into four subframes of 60 samples each. For every subframe, a 10th order Linear Prediction Coder (LPC) filter is computed using the unprocessed input signal. The LPC filter for the last subframe is quantized using a Predictive Split Vector Quantizer (PSVQ). The unquantized LPC coefficients are used to construct the short-term perceptual weighting filter, which is used to filter the entire frame and to obtain the perceptually weighted speech signal[9].

For every two subframes (120 samples), the open loop pitch period, LOL, is computed using the weighted speech signal. This pitch estimation is performed on blocks of 120 samples. The pitch period is searched in the range from 18 to 142 samples[9].

From this point the speech is processed on a 60 samples per subframe basis.Using the estimated pitch period computed previously, a harmonic noise shaping filter is constructed. The combination of the LPC synthesis filter, the formant perceptual weighting filter, and the harmonic noise shaping filter is used to create an impulse response. The impulse response is then used for further computations[9].

Using the pitch period estimation, LOL, and the impulse response, a closed loop pitch predictor is computed. A fifth order pitch predictor is used. The pitch period is computed as a small differential value around the open loop pitch estimate. The contribution of the pitch predictor is then subtracted from the initial target vector. Both the pitch period and the differential value are transmitted to the decoder[9].

Finally the non-periodic component of the excitation is approximated. For the high bit rate, Multipulse Maximum Likelihood Quantization (MP-MLQ) excitation is used, and for the low bit rate, an algebraic-code-excitation (ACELP) is used[9].

Parameters coded	Subframe 0	Subframe 1	Subframe 2	Subframe 3	Total
LPC indices					24
Adaptive codebook lags	7	2	7	2	18
All the gains combined	12	12	12	12	48
Pulse positions	20	18	20	18	73 (Note)
Pulse signs	6	5	6	5	22
Grid index	1	1	1	1	4
Total			189		

Table 2/G.723.1: Bit allocation of the 6.3 kbit/s coding algorithm

Parameters coded	Subframe 0	Subframe 1	Subframe 2	Subframe 3	Total
LPC indices					24
Adaptive codebook lags	7	2	7	2	18
All the gains combined	12	12	12	12	48
Pulse positions	12	12	12	12	48
Pulse signs	4	4	4	4	16
Grid index	1	1	1	1	4
Total					158

Table 3/G.723.1: Bit allocation of the 5.3 kbit/s coding algorithm

IV. G.722.2

The G.722.2 coder has input blocks of 320 speech samples in 16-bit uniform PCM format to encoded blocks of 132, 177, 253, 285, 317, 365, 397, 461 and 477 bits and from encoded blocks of 132, 177, 253, 285, 317, 365, 397, 461 and 477 bits to output blocks of 320 reconstructed speech samples. The sampling rate is 16 000 samples/s leading to a bit rate for the encoded bit stream of 6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 or 23.85 kbit/s. The coding scheme for the multi-rate coding modes is the so-called Algebraic Code Excited Linear Prediction Coder, hereafter referred to as ACELP. The multi-rate wideband ACELP coder is referred to as AMR-WB. The codec described in this Recommendation also utilizes an integrated Voice Activity Detector (VAD). The AMR-WB codec consists of nine source codecs with bit rates of 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85 and 6.60 kbit/s[10].

The codec is based on the code excited linear prediction (CELP) coding model. The input signal is pre-emphasized using the filter Hpre-emph(z) = $1-\mu z^{-1}$. The CELP model is then applied to the preemphasizedsignal[10].

In the CELP speech synthesis model model, the excitation signal at the input of the short-term LP synthesis filter is constructed by adding two excitation vectors from adaptive and fixed (innovative) codebooks. The speech is synthesized by feeding the two properly chosen vectors from these codebooks through the short-term synthesis filter. The optimum excitation sequence in a codebook is chosen using an analysis-by-synthesis search procedure[10].

Table 3/G.722.2 – Bit allocation of the AMR-WB coding algorithm for 20-ms frame

Mode	Parameter	1st subframe	2nd subframe	3rd subframe	4th subframe	Total per frame
23.85 kbit/s	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	88	88	88	88	352
	Codebook gain	7	7	7	7	28
	HB-energy	4	4	4	4	16
	Total					477

V. Conclusion

In this paper, We have Provided comparison between three different ITU – T based speech codec namely G.729, G.723.1 and G.722.2.The common thing in these coder have they provided speech codding of analogous signal by method other than PCM. The G.729 coder has algorithm for the coding of speech signals at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP). This coder is designed to operate with a digital signal obtained by first performing telephone bandwidth filtering of the analogue input signal, then sampling it at 8000 Hz. The G.723.1 is used for compressing the speech or other audio signal component of multimedia services at a very low bit rate as part of the overall H.324 family of standards. This coder has two bit rates associated with it: 5.3 and 6.3 kbit/s. The higher bit rate has greater quality. The lower bit rate gives good quality and provides system designers with additional flexibility. The G.722.2 is a high quality Adaptive Multi-Rate Wideband (AMR-WB) encoder and decoder that is primarily intended for 7 kHz bandwidth speech signals. AMR-WB operates at a multitude of bit rates ranging from 6.6 kbit/s to 23.85 kbit/s. The bit rate may be changed at any 20-ms frame boundary[10].

Reference

[1] RABINER (L.R.), SCHAFER (R.W.): Digital processing of speech signals, Prentice-Hall Int., 1978.

[2] ITAKURA (F.): Line spectral representation of linear predictive coefficients of speech signals, J. Acoust. Soc. Amer., Vol. 57, Supplement No. 1, S35, 1975.

[3] BISTRITZ (Y.), PELLERM (S.): Immittance Spectral Pairs (ISP) for speech encoding, Proc. ICASSP'93, pp. II-9 to II-12.

[4] PALIWAL (K.K.), ATAL (B.S.): Efficient vector quantization of LPC parameters at 24 bits/frame, IEEE Trans. Speech and Audio Processing, Vol. 1, No. 1, pp. 3-14, 1993.

[5] KABAL (P.), RAMACHANDRAN (R.P.): The computation of line spectral frequencies using Chebyshev polynomials, IEEE Trans. on ASSP, Vol. 34, No. 6, pp. 1419-1426, 1986.

[6] JÄRVINEN (K.), VAINIO (J.), KAPANEN (P.), HONKANEN (T.), HAAVISTO (P.), SALAMI (R.), LAFLAMME (C.), ADOUL (J.-P.): GSM enhanced full rate speech codec, Proc. ICASSP'97, pp. 771-774.

[7] HONKANEN (T.), VAINIO (J.), JÄRVINEN (K.), HAAVISTO (P.), SALAMI (R.), LAFLAMME (C.), ADOUL (J.-P.): Enhanced full rate speech codec for IS-136 digital cellular system, Proc. ICASSP'97, pp. 731-734.

[8] ITU – T (TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU, “Coding of speech at 8 kbit/s using conjugate structure algebraic-code-excited linear prediction (CS-ACELP)”

[9] ITU – T (TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU, “Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s”

[10] ITU – T (TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU, “Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)”