GSM-AMR
(GSM Adaptive Multi Rate codec)

Adaptive Multi-Rate (AMR) codec:

Speech and channel codec at gross bit-rates of

- 11.4 kbits/s (‘half-rate’) and
- 22.8 kbits/s (‘full-rate’).

In addition, the codec may operate at various combinations of speech and channel coding (codec mode) bit-rates for each channel mode. Adjusts its bit-rate according to network load

Speech coding rates:

- 12.2, 10.2, 7.95, 6.7, 5.9, 5.15, 4.75kb/ s

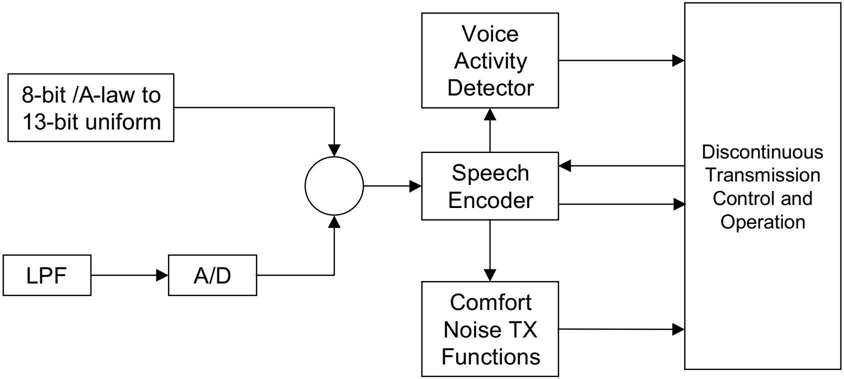

Figure 1:GSM-AMR audio-processing overview

- Based on CELP with 20 ms frame and 5 ms subframe
- Multirate- ACELP with 10th order short- term LPC and perceptual weighting (uses levinson)
- Encodes LSPs using split VQ
- An open loop LTP is first obtained and refined by closed loop
- Highest bit rate provides toll quality & half rate provides communications quality

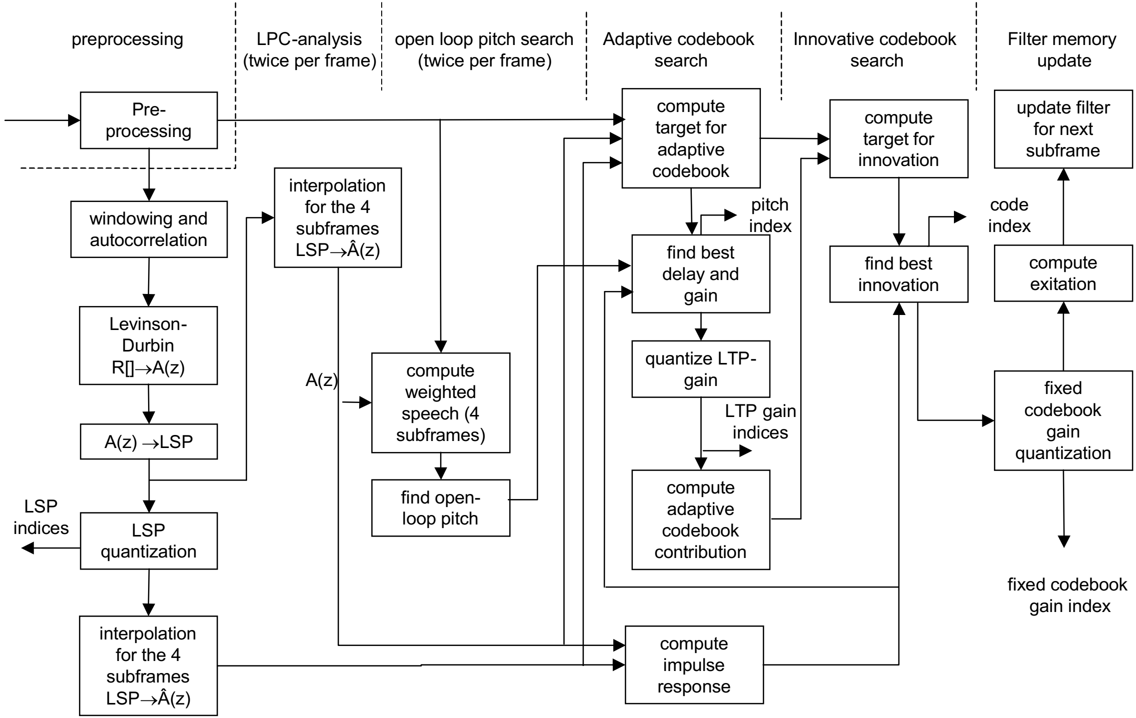

Figure 2: GSM-AMR speech coder

Principles of the GSM adaptive multi-rate speech encoder

The AMR codec uses eight source codecs with bit-rates of 12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15 and 4.75 kbit/s.
The codec is based on the code-excited linear predictive (CELP) coding model. A 10th order linear prediction (LP), or short-term, synthesis filter is used (=10 (quantified) linear prediction (LP) parameters).
A long-term, or pitch, synthesis filter is given where T is the pitch delay and g_p is the pitch gain. The pitch synthesis filter is implemented using the so-called adaptive codebook approach.
In the CELP speech synthesis 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 in which the error between the original and synthesized speech is minimized according to a perceptually weighted distortion measure.
The coder operates on speech frames of 20 ms corresponding to 160 samples at the sampling frequency of 8000 sample/s. At each 160 speech samples, the speech signal is analysed to extract the parameters of the CELP model (LP filter coefficients, adaptive and fixed codebooks' indices and gains). These parameters are encoded and transmitted. At the decoder, these parameters are decoded and speech is synthesized by filtering the reconstructed excitation signal through the LP synthesis filter.
LP analysis is performed twice per frame for the 12.2 kbit/s mode and once for the other modes. For the 12.2 kbit/s mode, the two sets of LP parameters are converted to line spectrum pairs (LSP) and jointly quantized using split matrix quantization (SMQ) with 38 bits. For the other modes, the single set of LP parameters is converted to line spectrum pairs (LSP) and vector quantized using split vector quantization (SVQ).
The speech frame is divided into 4 subframes of 5 ms each (40 samples). The adaptive and fixed codebook parameters are transmitted every subframe. The quantized and unquantized LP parameters or their interpolated versions are used depending on the subframe. An open-loop pitch

lag is estimated in every other subframe (except for the 5.15 and 4.75 kbit/s modes for which it is done once per frame) based on the perceptually weighted speech signal.

Then the following operations are repeated for each subframe:

- The target signal x(n) is computed by filtering the LP residual through the weighted synthesis filter W(z)H(z) with the initial states of the filters having been updated by filtering the error between LP residual and excitation (this is equivalent to the common approach of subtracting the zero input response of the weighted synthesis filter from the weighted speech signal).
- The impulse response, h(n) of the weighted synthesis filter is computed.
- Closed-loop pitch analysis is then performed (to find the pitch lag and gain), using the target x(n) and impulse response h(n), by searching around the open-loop pitch lag. Fractional pitch with 1/6th or 1/3rd of a sample resolution (depending on the mode) is used.
- The target signal x(n) is updated by removing the adaptive codebook contribution (filtered adaptive codevector), and this new target, $x_2(n)$, is used in the fixed algebraic codebook search (to find the optimum innovation).
- The gains of the adaptive and fixed codebook are scalar quantified with 4 and 5 bits respectively or vector quantified with 6-7 bits (with moving average (MA) prediction applied to the fixed codebook gain).
- Finally, the filter memories are updated (using the determined excitation signal) for finding the target signal in the next subframe.

Sample bit allocation of the AMR codec modes is shown in table 1. In each 20 ms speech frame, 95, 103, 118, 134, 148, 159, 204 or 244 bits are produced, corresponding to a bit-rate of 4.75, 5.15, 5.90, 6.70, 7.40, 7.95, 10.2 or 12.2 kbit/s.

Example of Bit allocations:

Table 1: Bit allocation of the AMR coding algorithm for 20 ms frame

Mode	**Parameter**	**1st subframe**	**2nd subframe**	**3rd subframe**	**4th subframe**	**total per frame**
12.2 kbits/s (GSM EFR)	2 LSP sets					38
	Pitch delay	9	6	9	6	30
	Pitch gain	4	4	4	4	16
	Algebraic code	35	35	35	35	140
	Codebook gain	5	5	5	5	20
	TOTAL					**244**
4.75 kbits/s	LSP sets					23
	Pitch delay	8	4	4	4	20
	Algebraic code	9	9	9	9	36
	Gains	8		8		16
	TOTAL					**95**