## 5.4 Switching of Coding Modes

### 5.4.1 General description

As described in subclause 5.2 and 5.3, the EVS codec supports a CELP coding mode as well as a MDCT coding mode. The transitions between both within the same bit rate and audio bandwidth are described in 5.4.2 and 5.4.3.

The CELP or LP-based coding mode can operate on different sample rates depending on the frame configuration. The procedure how to handle sample rate changes during the encoding process is described in 5.4.4.

The switching between primary and AMR-WB IO modes is described in 5.4.5.

The handling of transitions in the context bit rate switches is described in 5.4.6.

### 5.4.2 MDCT coding mode to CELP coding mode

When a CELP encoded frame is preceded by a MDCT based encoded frame, the memories of the CELP encoded frame have to be updated before starting the encoding of the CELP frame. These memories include:

* Adaptive codebook memory
* LPC synthesis filter memory
* Weighting filter denominator memory (used to compute the target signal)
* The factor  of the tilt part of the innovative codebook pre-filter
* De-emphasis filter memory
* MA/AR prediction memories used in end-frame LSF quantization
* Previous quantized end-frame LSP (for quantized LPC interpolation)
* Previous quantized end-frame LSF (for mid-frame LSF quantization)

The CELP memories update is performed depending on the bitrate and the previous encoding mode (either MDCT based TCX or HQ MDCT). In general, three different MDCT to CELP (MC1-3) transition methods are supported. The following table 149 lists the different cases depending on MDCT mode and bit rate.

Table 149: MDCT to CELP transition modes

|  |  |  |  |
| --- | --- | --- | --- |
| Switching from | Switching to | Bitrate (kbps) | Transition mode |
| HQ MDCT | CELP | 7.2 | MC1 |
| HQ MDCT | CELP | 8 | MC1 |
| TCX | CELP | 9.6 | MC2 |
| TCX or HQ MDCT | CELP | 13.2 | MC1 |
| TCX | CELP | 16.4 | MC2 |
| HQ MDCT | CELP | 16.4 | MC3 |
| TCX | CELP | 24.4 | MC2 |
| HQ MDCT | CELP | 24.4 | MC3 |
| TCX or HQ MDCT | CELP | 32 | MC1 |
| HQ MDCT | CELP | 64 | MC1 |

In following subclauses, the MDCT to CELP transitions are described in detail. Note that this description only considers switching cases within the same bit rate.

#### 5.4.2.1 MDCT to CELP transition 1 (MC1)

MC1 is used when the previous frame was coded with HQ MDCT and the current frame is coded with CELP. In this case, the CELP state variables are reset in the current frame to predetermined (fixed) values. In particular the following memories are reset to 0 in the CELP encoder:

* Resampling memories of the CELP synthesis signal
* Pre-emphasis and de-emphasis memories
* LPC synthesis memories
* Past excitation (adaptive codebook memory)

The old LPC coefficients and associated representations (LSP, LSF) and CELP gain quantization memories are reset to predetermined (fixed) values. Since the past excitation is not available, the CELP coder in the current frame is forced to operate in Transition coding (TC) , i.e. without any adaptive codebook. The LPC coefficients from the previous frame are not available, therefore only one set of LPC coefficients corresponding to the end of frame are coded and used for all subframes of the current frame.

#### 5.4.2.2 MDCT to CELP transition 2 (MC2)

MC2 is designed for CELP transitions coming from the MDCT based TCX mode. The TCX shares the same LPC analysis and quantization as in CELP, as described in subclause 5.3.3.2.1. The MA/AR/LSP/LSF memories are consequently updated during MDCT based TCX encoding, as it is done in CELP encoding. The only exception is at 9.6kbps, where the weighted LPC are quantized (instead of the unweighted LPC as in CELP) as described in subclause 5.3.3.2.1.1.1. In that case, the MA/AR/LSP/LSF memories are re-computed in the unweighted domain as described in subclause 5.3.3.2.1.1.2.

Moreover, MDCT based TCX includes an internal decoder which generates a decoded time-domain signal at the CELP sampling rate as described in subclause 5.3.3.2.12.1. This signal and the quantized LPC are then used to update CELP memories as described in 5.3.3.2.12.2, including the adaptive codebook memory, the LPC synthesis filter memory, the weighting filter denominator memory and the de-emphasis filter memory.

#### 5.4.2.3 MDCT to CELP transition 3 (MC3)

Similarly to MC1, the CELP state variables are reset to predetermined values (in general 0), with the following exceptions::

* The factor  is set to 0.3.
* The LSF quantization is run in safety-net mode, such that no prediction is used and the MA/AR memories are not used.
* The previous quantized end-frame LSP/LSF and the quantized mid-frame LSP/LSF are set to the current quantized end-frame LSP/LSF.

### 5.4.3 CELP coding mode to MDCT coding mode

When a MDCT encoded frame is preceded by a CELP encoded frame, a beginning portion of the MDCT encoded frame cannot be reconstructed properly due to the aliasing introduced by the missing previous MDCT encoded frame. To solve this problem, two approaches are used depending on the MDCT based coding mode (either MDCT based TCX or HQ MDCT). These approaches are described in detail in the following subclauses.

#### 5.4.3.1 CELP coding mode to MDCT based TCX coding mode

When the previous frame is CELP and the current frame is MDCT based TCX, the MDCT length is increased, the left folding point is moved towards the past and the left overlap length is reduced such that the current MDCT based TCX can reconstruct the whole 20ms frame, without the need for the previous (and missing in this case) MDCT based frame. This is illustrated in the figure below.

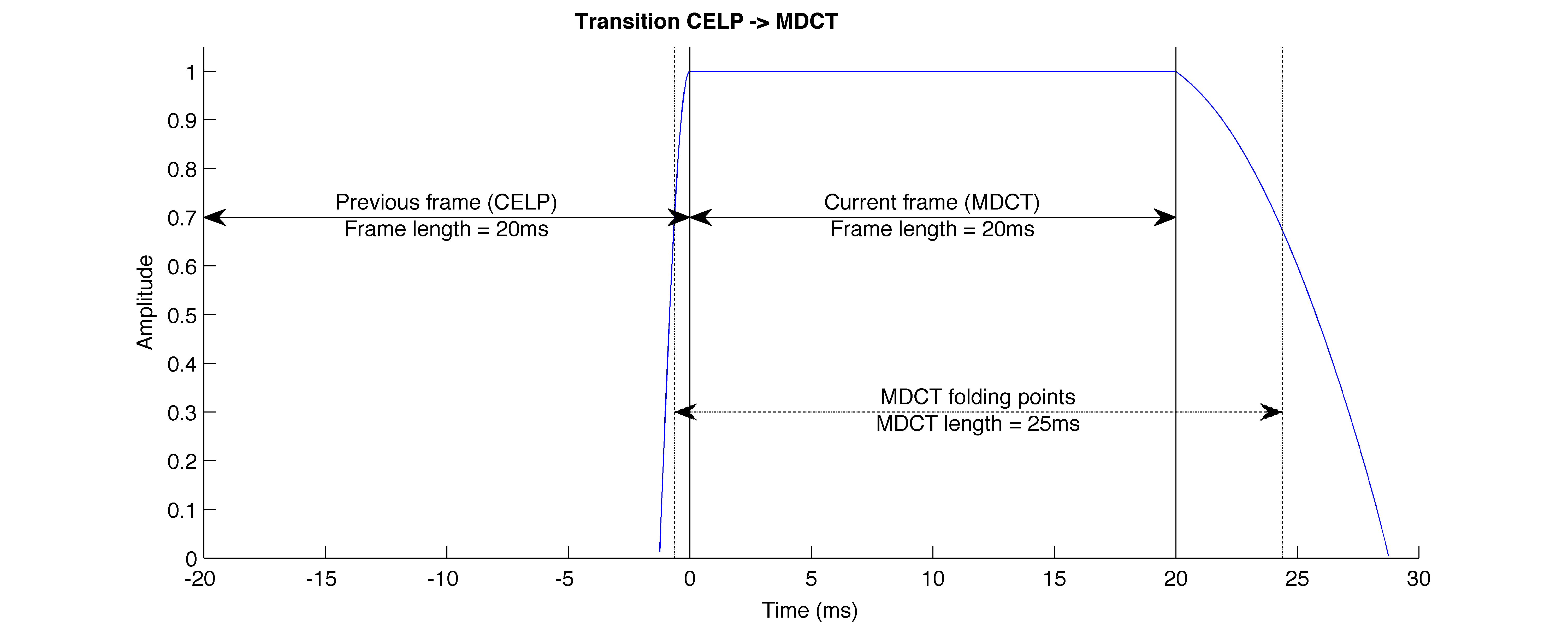


Figure 85: CELP to MDCT based TCX transition window (right part is here ALDO)

The right part of the transition window is not changed, such that it can be used in the next MDCT based frame as if it was a normal (non-transition) MDCT based frame. Similarly to the non-transition case, the right part of the transition window can have different shapes like ALDO, HALF or MINIMAL as described in subclause 5.3.2.3.

The left folding point is moved towards the past at 0.625ms before the transition border. This is equivalent to increasing the MDCT length from 20ms to 25ms. The corresponding numbers of MDCT bins are given in subclause 5.3.3.1.1.

The left part of the MDCT window is modified such that window segment with weight 1 covers the whole 20ms frame until the transition border, and the window segment before the transition border is a sine window with length 1.25ms

 (1284)

with  is the window length in samples and  is the sampling rate of the time-domain signal.

After windowing and MDCT, the MDCT-based TCX frame is encoded as described in subclause 5.3.3.

#### 5.4.3.2 CELP coding mode to HQ MDCT coding mode

When the previous frame is CELP and the current frame is to be coded by HQ MDCT, the current frame is a transition frame in which two types of coding are used:

* Constrained CELP coding and (when required) simplified time-domain BWE coding
* HQ MDCT coding with a modified window

Constrained CELP coding means here that CELP is restricted to cover only the first subframe of the current frame, to code only a subset of CELP parameters, and to reuse parameters (LPC coefficients) from the previous CELP frame. These constraints are set to minimize the bit budget taken by continuing CELP coding in the current transition frame, this bit budget being taken out of HQ MDCT coding.

##### 5.4.3.2.1 Constrained CELP coding and simplified BWE coding

The bit budget for CELP and BWE in the current (transition) frame is determined depending on the CELP coder used in the previous frame (12.8 kHz or 16 kHz) and coded audio bandwidth in the current frame. The following pseudo-code describes how this bit budget is subtracted from the total bit budget for the current frame (total\_budget):

num\_bits = total\_budget

if CELP 12.8kHz was used in the previous frame

cbrate = min(core\_bitrate, ACELP\_24k40)

if cbrate  ACELP\_11k60, num\_bits = num\_bits – 1, end

num\_bits = num\_bits – table\_ACB\_bits[cbrate,GENERIC]

num\_bits = num\_bits – table\_gain\_bits [cbrate,TRANSITION]

num\_bits = num\_bits – table\_FCB\_bits [cbrate,GENERIC]

else (CELP 16 kHz was used in the previous frame)

if core\_bitrate  ACELP\_8k00, cbrate = ACELP\_8k00

else if core\_bitrate  ACELP\_14k80, cbrate = ACELP\_14k80

else cbrate = min(core\_bitrate, ACELP\_22k60)

end

if cbrate  ACELP\_11k60, num\_bits = num\_bits – 1, end

num\_bits = num\_bits – table\_ACB\_bits\_16kHz[cbrate,GENERIC]

num\_bits = num\_bits – table\_gain\_bits\_16kHz [cbrate,GENERIC]

num\_bits = num\_bits – table\_FCB\_bits\_16kHz [cbrate,GENERIC]

end

if bandwidth is not NB and (bandwidth is not WB and CELP 16 kHz was not used in the previous frame)

num\_bits = num\_bits –(6+6)

end

The bit rate for CELP coding is any case saturated by the minimum of HQ MDCT coding bit-rate and a predetermined bit-rate value (24 kbit/s or 22.6 kbit/s depending on whether the CELP core is at 12.8 or 16 kHz), then the numbers of bits allocated to CELP coding is subtracted and the remaining bit budget (denoted ‘num\_bits’) is reserved for HQ MDCT coding normally operating at a bit rate ‘core\_bitrate’ in the current frame. CELP coding in the extra subframe is configured to operate as if the current frame was CELP at a bit-rate denoted ‘cbrate’; this CELP bit-rate depends on the CELP coder used in the previous frame (12.8 kHz or 16 kHz).

The coded CELP parameters in this extra subframe are: pitch filter flag (1 bit) if the CELP bit-rate is  11.6 kbit/s, pitch index for the adaptive codebook (ACB), codebook gains, fixed codebook (FCB) index. The bit allocation tables from CELP coding in Generic or Transition coding at 12.8 and 16 kHz are reused.

Besides, 12 bits are used for BWE in the extra subframe to code one gain (6 bits) and one pitch index (6 bits) for the high band above CELP synthesis.

Note that the current frame being a transition frame, one bit is used to indicate the type of CELP coding (12.8 kHz or 16 kHz) used in the extra CELP subframe; this bit is necessary to be able to decode correctly the transition frame in case of frame erasures.

LPC coefficients from the end of the previous frame are reused to code the extra subframe; constrained CELP coding reuses the subframe excitation coding with the same CELP core coder (12.8 kHz or 16 kHz) as in the previous frame, and this subframe coding is adapted from the procedure described in clause 5.2.3.1.

When the coded audio bandwidth is higher than the bandwidth of the core CELP coder, simplified BWE coding is applied. The previous and current input frames are high-pass FIR filtered to obtain the high-band; the cutoff frequency (6.4 or 8 kHz) depends on the core CELP coder. Then, pitch search based on correlation in the high band provides an estimated pitch lag and gain which are coded with 6 bits each.

##### 5.4.3.2.2 HQ MDCT coding with a modified analysis window

HQ MDCT coding in the transition frame is identical to clause 5.3.4, except the MDCT analysis window is modified and the bit budget for HQ MDCT coding in the current frame is decreased as described in clause 5.4.3.2.1.



Figure 85a: Modified MDCT window in transition frame (CELP to MDCT transition)

The modified MDCT window is designed to avoid aliasing in the first part of the frame as shown in Figure 85a. Its shape also allows cross-fading between the synthesis from constrained CELP and simplified BWE and the synthesis from HQ MDCT as described in clause 6.3.3.2.3. Note that the frames labeled CELP and MDCT in Figure 85a represent the new frames of signal (20 ms) entering in the encoder; the actual coded frame is delayed by the encoder lookahead.

### 5.4.4 Internal sampling rate switching

The LP-based coding within EVS operates at two internal sampling rates, 12.8 kHz and 16 kHz. In active frames the 12.8 kHz internal sampling is employed at lower bit-rates (≤ 13.2 kbps) while the 16 kHz internal sampling is employed at higher bit-rates (≥ 16.4 kbps). Further in LP-based CNG, the 12.8 kHz internal sampling is employed at bit-rates ≤ 8.0 kbps while 16 kHz internal sampling is employed at bit-rates ≥ 9.6 kbps. Consequently a CELP internal sampling rate switching can happen either 1) in case of bit-rate switching or 2) in case of switching between active segments and LP-based CNG segments at 9.6 kbps and 13.2 kbps.

The MDCT-based TCX operates at 4 different internal sampling rates, 12.8, 16, 25.6 and 32 kHz. MDCT-based TCX internal sampling rate corresponds to the rate used for computing and transmitting its LP filter, filter employed for shaping the quantization noise in frequency domain. The same internal sampling rate is used for generating the low rate decoded signal computed at both encoder and decoder sides for updating memories of an eventual next CELP frame. The sampling-rate switching in MDCT-Based TCX can only happen either 1) in case of bit-rate switching or 2) in case of switching from CELP at 13.2kbps or switching from LP-based CNG segments at 9.6 kbps.

When changing the internal sampling rate, a number of memory and buffer updates needs to be done. These are described in subsequent subclauses.

#### 5.4.4.1 Reset of LPC memory

In case of internal sampling rate switching, the LSF quantization is run in safety-net mode, such that no prediction is used and the MA/AR memories are not used.

#### 5.4.4.2 Conversion of LP filter between 12.8 and 16 kHz internal sampling rates

When switching between internal sampling rates of 12.8 kHz and 16 kHz, the previous LP filter needs to be converted both at the encoder and the decoder between these two sampling rates in order to determine the interpolated LP parameters of the current frame. For this purpose, the LP filter of the previous frame could be recomputed at the current sampling rate based on the past synthesis signal that is already available. However, this would require complete LP analysis and resampling the past synthesis signal both at the encoder and the decoder. A less complex method is used here based on re-estimating the LP filter from its power spectrum modified corresponding to the current sampling frequency. The autocorrelation is computed from this modified power spectrum for solving the parameters of the converted LP filter with the Levinson Durbin algorithm. The converted LP filter is finally transformed to its line spectrum frequency representation for interpolation with the corresponding parameters of the current frame.

The computation and modification of the power spectrum as well as the computation of the autocorrelation are described in the following subclauses. The Levinson-Durbin algorithm is described in subclause 5.1.9.4 and the determination of the line spectrum frequencies in subclause 5.1.9.5.

Note that only the quantized LP filter is converted. Although the perceptual weighting filter uses the unquantized LP filter at the encoder, it is sufficient to use the converted quantized LP filter for interpolation when switching between internal sampling rates. This approximation avoids an additional conversion procedure for the unquantized LP filter.

##### 5.5.4.1.1 Modification of the Power Spectrum

When switching the internal sampling rate down to 12.8 kHz from 16 kHz, the converted LP filter models the power spectrum of the LP filter originally estimated at a sampling rate of 16 kHz up to the new cut-off frequency. This is accomplished by computing the power spectrum of the LP filter at  frequency points equispaced in , corresponding to the Nyquist frequency at 12.8 kHz sampling rate. This frequency range of the power spectrum is then mapped onto  when computing the autocorrelation for solving the parameters of the converted LP filter.

Conversely when switching the internal sampling rate up to 16 kHz from 12.8 kHz, the power spectrum of the LP filter originally estimated at a sampling rate of 12.8 kHz is computed at  frequencies equispaced in . This frequency range of the power spectrum is mapped onto  for autocorrelation computation. The power spectrum unknown at frequencies  is approximated by extending the power spectrum value at  over this range by  values for autocorrelation computation. This procedure thus re-estimates the LP filter at sampling frequency 16 kHz with an approximated, extended upper band.

By choosing , all power spectrum and autocorrelation computations can be accomplished on two equispaced frequency grids, one including the frequency points  and one the points . Converting down to 12.8 kHz is hence equivalent to dropping 10 last values away from the power spectrum of the original LP filter. Similarly, converting up to 16 kHz translates to adding 10 approximated values to the power spectrum of the original LP filter.

The same procedure is applied identically both at the encoder and the decoder.

##### 5.5.4.1.2 Computation of the Power Spectrum

The LP filter is converted to another sampling rate by truncating or extending its power spectrum

, , ()

to the frequency range that corresponds to the new sampling rate and re-estimating linear prediction coefficients from the autocorrelation computed from this modified spectrum. For reduced complexity, the power spectrum is expressed on the real axis by utilizing the line spectrum frequency decomposition

, ()

where the polynomials  and  are defined as in subclause 5.1.9.5. The zeros of these two polynomials give the line spectrum frequencies of .

Because of the symmetry properties of the polynomials  and , they can be expressed as the polynomialsand of . It can be shown that for an LP filter of an even order,

, . ()

The use of this expression is motivated by the observation that the polynomials and can be evaluated efficiently for a given  with Horner’s method [34]. The value of these polynomials is needed on frequency grids described in subclause 5.5.4.1.1. The expression (1287) obtains a particularly simple form at  that can be utilized in computation.

The polynomials and  can be derived by substituting the explicit forms of the Chebyshev polynomials

,     ()

to the Chebyshev series representation of  and [34]. This same representation is employed in subclause 5.1.9.5 for the determination of the line spectrum frequencies. The explicit forms of  can readily be written by using the recursion

 ()

By definition the zeros of and are respectively the cosines of the even and odd line spectrum frequencies. The coefficients of these polynomials are thus readily obtained when the line spectrum frequencies of are known. Given that the order of  is 16, the polynomial is of order 8 and can be expressed as

 ()

The leading coefficient  is constant. This relation yields a simple recursion for solving the coefficients of from the even line spectrum pairs  for . The coefficients of  are obtained correspondingly from the odd line spectrum pairs*.*

If no line spectrum frequency representation is available for, one can alternatively first compute the coefficients of the polynomials and  from those of and then solve the coefficients of and from a set of equations that relate the two representations. These relating equations can be derived by substituting the explicit forms of the Chebyshev polynomials to the Chebyshev series representation of  and . This approach is employed when switching from the AMR-WB IO mode, which uses the immittance spectrum frequency representation instead of line spectrum frequencies.

##### 5.5.4.1.3 Computation of the Autocorrelation

The autocorrelation of the LP filter is obtained by the inverse Fourier Transform of the modified power spectrum. Since the power spectrum is real and symmetric, the relation between the autocorrelation and the power spectrum can be expressed through the integral

,           *k* = 0, 1, … ()

Because the power spectrum is real and symmetric, , it suffices to evaluate the integral over only the upper half of the unit circle. Due to this symmetry, the rectangle rule for approximating the integral can be expressed as

,       *k* = 0, 1, … ()

where  is the set of frequencies equispaced in [0, **], but excluding 0 and **to avoid double counting. The number of these frequencies is hereafter assumed odd.

Note that in sequel the factor  is omitted for simplicity from the approximation of the autocorrelation. Namely, the autocorrelation can be scaled for convenience, because the resulting linear prediction coefficients are invariant to this scaling.

The expression of autocorrelation can be rewritten equivalently for more efficient computation by utilizing the symmetries of the cosine term relative to through the following trigonometric identities:

 ()

where *k*  0, 1, … and . The operator  rounds to the nearest integers towards minus infinity. The term  simply generates the sequence 2, 0, 2, 0, 2, 0, 2, 0, 2, … and is hence readily implementable.

By employing the two trigonometric identities given above, the autocorrelation can be rewritten as

 ()

where  is the set of  frequencies equispaced in but excluding 0 and **. The cosine term of the autocorrelation is evaluated using the recursion

 ()

starting from  and . The value of  is stored in a table that holds all the entries needed for .

When switching the internal sampling rate from 16 kHz down to 12.8 kHz, a grid of  equispaced frequency points is used. Switching from 12.8 kHz up to 16 kHz uses a grid of points, see subclause 5.5.4.1.1.

#### 5.4.4.3 Extrapolation of LP filter

In case of sampling rate switching involving at least a sampling rate being neither 12.8 nor 16 kHz, the previous LP filter is not converted. Instead, the previous quantized end-frame LSP/LSF and the quantized mid-frame LSP/LSF are set to the current quantized end-frame LSP/LSF.

The LP filter is also extrapolated when the conversion of LP filter between 12.8 and 18 kHz, described in suclause 5.4.4.2, does not produce a stable filter. The stability of the filter is detected during the Levinson Durbin algorithm described in subclause 5.1.9.4. A filter is detected as unstable when at least one of the reflection coefficients has an absolute value greater than 0.99945.

#### 5.4.4.4 Buffer resampling with linear interpolation

Switching the internal sampling rate requires several memories to be resampled. . In order to reduce the complexity of the resampling processing, a simple linear interpolation is used most of the time instead of a conventional low-pass filtering method.

The basic operation for interpolating a point is done as follows:

 ()

where  is the new resampled buffer of size  and  is the old buffer of size . The index  is initialized to 0 while index  is equal to the integer part of the position , . The position is initialized to:

 ()

where  is the increment of the position  for each unit increment of the index n.

#### 5.4.4.5 Update of CELP input signal memories

The pre-emphasized input signal  defined in subclause 5.1.4 is updated at both CELP internal sampling rate 12.8 and 16 kHz at any bit-rates. No specific processing is then needed in case of sampling rate switching.

The weighed synthesis filterstate is updated in three different ways in case of sampling rate switching:

* It is set to zero if a sampling rate different from 12.8 and 16 kHz is involved.
* At bit-rates <= 8, 13.2, 32 and 64 kbps, the memory states is updated in previous frame as usual and the difference in sampling rate between the previous and the current frame is simply ignored.
* Otherwise, the state is recomputed by filtering the resampled LPC synthesis filter state obtained in subclause 5.4.4.7 through the filter filter and by taking computing the error between the obtained signal and the input weighted signal. The memory state is used for computed the target signal of the next frame as defined in subclause 5.2.3.1.2.

#### 5.4.4.6 Update of MDCT-based TCX input signal memories

The past of the input signal and the past of weighted signal are needed for MDCT-based TCX and both past signal are resampled as in sub-clause 5.4.4.4.

#### 5.4.4.7 Update of CELP synthesis memories

For being able to make a seamless transition from CELP to CELP or from MDCT-based TCX to CELP, the following memory states have to be maintained:

* The adaptive codebook state
* The LPC synthesis filter state
* The de-emphasis state

The three memory states are maintained at both encoder and decoder side in CELP mode and in MDCT-based TCX coding mode. The following specific processing is performed in case of internal sampling rate switching.

The adaptive codebook state covers at the encoder side a frame, i.e. 20 ms. In case of internal sampling rate switching between 12.8 and 16 kHz, the adaptive codebook is resampled with the method described in subclause 5.4.4.4. If it involves at least a sampling rate different from 12.8 and 16 kHz, the adaptive codebook is reset with zeros.

The LPC synthesis filter state doesn’t cover a fixed time duration but a fixed number samples equal to the order of the LPC. This order is always 16. For being able to resample this state at any of the sampling rate between 12.8 and 48 kHz, the memory of the LPC synthesis filter state is extended from 16 to 60 samples, which represents 1.25ms at 48 kHz. The memory resampling from sampling rate  Hz to sampling rate Hz can summarized as:

 ()

where  is the function resampling the input buffer *x* from ** to ** samples as described in subclause 5.4.4.4. *L\_SYN\_MEM* is the largest size in samples that the memory can cover and is equal to 60 samples. At any sampling rate and at any time, *mem\_syn\_r* is updated with the last *L\_SYN\_MEM* output samples and is then eventually resampled in case of internal sampling rate switching at the beginning of the next frame.

The de-emphasis has a fixed order of 1, which represents also a different time duration at different sampling rates. However the resampling stage is not performed and the memory update in done as usual even in case of intern sampling rate switching.

### 5.4.5 EVS primary and AMR-WB IO

The codec support a seamless switching between primary and AMR-WB IO modes. While most of memories and past buffers are shared between the two modes, there are some particularities that need to be properly handled. The following scenarios can happen:

#### 5.4.5.1 Switching from primary modes to AMR-WB IO

When a CELP based AMR-WB IO encoded frame is preceded by a primary mode encoded frame, the memories of the CELP AMR-WB IO encoded frame have to be updated or converted before starting the encoding. These include:

* Convert previous quantized end-frame LSFs to ISFs
* Convert previous quantized end-frame LSPs to ISPs
* Set previous un-quantized end-frame ISPs to converted quantized end-frame ISPs
* Convert previous CNG quantized end-frame LSPs to ISPs
* Limit index of last encoded CNG energy to 63
* Reset the gain quantization memory to -14.0.
* In case the switching happens in the SNG segment, force SID frame
* Reset AR model LP quantizer memory

In case the AMR-WB IO frame is preceded by CELP primary frame at 16 kHz internal sampling rate, the processing described in subclause 5.4.4 is performed.

Finally in case the AMR-WB IO frame is preceded by MDCT primary frame, the processing described in subclause 5.4.2 is performed.

#### 5.4.5.2 Switching from AMR-WB IO mode to primary modes

When a primary mode encoded frame is preceded by a CELP based AMR-WB IO encoded frame, the memories of the primary mode encoded frame have to be updated or converted before starting the encoding. These include:

* First three ACELP frames are processed using safety-net LP quantizer
* Convert previous quantized end-frame ISFs to LSFs
* Convert previous quantized end-frame ISPs to LSPs
* Convert previous CNG quantized end-frame LSPs to ISPs
* Reset BWE past buffers
* reset the unvoiced/audio signal improvement memories

In case of CELP at 16 kHz internal sampling rate primary mode frame is preceded by the AMR-WB IO frame, the processing described in subclause 5.4.4 is performed.

Finally in case the MDCT primary frame is preceded by AMR-WB IO frame, the processing described in subclause 5.4.3 is performed.

### 5.4.6 Rate switching

A seamless switching between all EVS primary rates is supported in the codec. Since most of the states and memories are shared and maintained at any bit-rates, a complete re-initialization is not needed. The coding tools are able to be reconfigured at the beginning of any frame. The different bit-rate dependent setups of each tool are described in each corresponding subclause. Rate switching doesn’t require any specific handling, except in the following scenarios.

#### 5.4.6.1 Rate switching along with internal sampling rate switching

In case the internal sampling rate changes when switching the bit-rate, the processing described in subclause 5.4.4 is performed at first.

#### 5.4.6.2 Rate switching along with coding mode switching

In case the internal sampling rate changes when switching the bit-rate, the processing described in subclause 5.4.3 is performed. New possible transitions from MDCT to CELP mode are possible during rate switching compared to table 149. table 150 lists all different cases achievable during rate switching depending on MDCT mode and the new CELP bit rate.

Table 150: MDCT to CELP transition modes in case of rate switching

|  |  |  |  |
| --- | --- | --- | --- |
| Switching from | Switching to | CELP Bitrate (kbps) | Transition mode |
| HQ MDCT or TCX | CELP | 7.2 | MC1 |
| HQ MDCT or TCX | CELP | 8 | MC1 |
| TCX | CELP | 9.6 | MC2 |
| HQ MDCT | CELP | 9.6 | MC3 |
| HQ MDCT or TCX | CELP | 13.2 | MC1 |
| TCX | CELP | 16.4 | MC2 |
| HQ MDCT | CELP | 16.4 | MC3 |
| TCX | CELP | 24.4 | MC2 |
| HQ MDCT | CELP | 24.4 | MC3 |
| HQ MDCT or TCX | CELP | 32 | MC1 |
| HQ MDCT or TCX | CELP | 64 | MC1 |

If the internal sampling rate is also changing, the processing of subclause 5.4.4 is performed beforehand.

## 5.5 Frame erasure concealment side information

The codec has been designed with emphasis on performance in frame erasure conditions and several techniques limiting the frame erasure propagation have been implemented, namely the TC mode, the safety-net approach for LSF quantization, and the memory-less gain quantization. To further enhance the performance in frame erasure conditions, side information, consisting of concealment/recovery parameters, is sent in the bitstream to the decoder. This supplementary information improves the frame erasure concealment (FEC) and the convergence and recovery of the decoder after erased frames. The detailed concealment and recovery processing is described in [6].

The concealment/recovery parameters that are transmitted to the decoder depend on the bitrate and coding mode. They will be described in the following subclauses together with the information about configurations when they are transmitted.

### 5.5.1 Signal classification parameter

The signal classification parameter is determined based on the classification for FEC, described in subclause 5.1.13.3. The classification uses the following five classes to classify speech signals: UNVOICED, UNVOICED TRANSITION, VOICED TRANSITION, ONSET and VOICED. For the purpose of the FEC classification, inactive signals fall into the UNVOICED category. Though there are five signal classes, they can be encoded with only two bits as the differentiation of the both TRANSITION classes can be done unambiguously determined based on the class of the preceding frame. The rules for the FEC signal classification are described in subclause 5.1.13.3.3.

The signal classification parameter is not transmitted at lowest bitrates. Further, it does not need to be transmitted in coding modes that allow to classify the frame implicitly, e.g. in the UC and VC modes. The signal classification parameter is transmitted in GC and TC modes at 13.2 kb/s, 32 kb/s and 64 kb/s.

### 5.5.2 Energy information

Precise control of the speech energy is very important in frame erasure concealment. The importance of the energy control becomes more evident when a normal operation is resumed after an erased block of frames. Since VC and GC modes are heavily dependent on prediction, the actual energy cannot be properly estimated at the decoder. In voiced speech segments, the incorrect energy can persist for several consecutive frames, which can be very annoying, especially when this incorrect-valued energy increases.

To better control the energy of the synthesized sound signal at the decoder in case of frame erasure, the energy information is estimated and sent using 5 bits. The goal of the energy control is to minimize energy discontinuities by scaling the synthesized signal to render the energy of the signal at the beginning of the recovery frame (a first non erased frame received following frame erasure) to be similar to the energy of the synthesized signal at the end of the last frame erased during the frame erasure. The energy of the synthesized signal in the received first non erased frame is further made converging to the energy corresponding to the received energy parameter toward the end of that frame while limiting an increase in energy.

The energy information is the maximum of the signal energy for frames classified as VOICED or ONSET, or the average energy per sample for all other frames. For VOICED or ONSET frames, the maximum signal energy is computed pitch-synchronously at the end of the current frame as follows:

 ()

where  is the frame length for the 12.8 kHz internal sampling rate, and  is the frame length for the 16 kHz sampling rate. Signal  is the local synthesis signal sampled at 12.8 kHz or 16 kHz depending on the internal sampling rate. The integer pitch period length is the rounded pitch period of the last subframe, i.e.  for the 12.8 kHz core, and  for the 16 kHz core.

For all other classes, is the average energy per sample of the last two subframes of the current frame, i.e.,

 ()

The energy information is quantized using a 5-bit linear quantizer in the range of 0 dB to 96 dB with a step of 3 dB. The quantization index is given by

 ()

The index is limited to the range [0,…, 31].

The energy information is sent only in GC mode at 32 and 64 kb/s.

### 5.5.3 Phase control information

The phase control is particularly important when recovering after a lost voiced segment of a signal. After a block of erased frames, the decoder memories become unsynchronized with the encoder memories. Sending some phase information helps in the re-synchronization of the decoder. The rough position of the last glottal pulse in the previous frame is sent.

Let  be the integer closed-loop pitch lag for the last subframe of the previous frame. The position of the last glottal pulse, , is searched among the last samples of the previous frame by looking for the sample with the maximum amplitude of low-pass filtered LP residual signal. A simple FIR low-pass filter with coefficients 0.25, 0.5 and 0.25 is used.

The position of the last glottal pulse, , is encoded using 8 bits in the following manner. The precision used to encode the position of the pulse depends on the integer part of the closed-loop pitch lag for the first subframe of the current frame, . This is possible because this value is known both at the encoder and the decoder, and is not subject to erasure propagation after one or several frame losses. When is less than 128, the position of the last glottal pulse, relative to the end of the previous frame, is encoded directly with a precision of one sample. When , the position of the last glottal pulse, relative to the end of the previous frame, is encoded with a precision of two samples by using a simple integer division, i.e., . Finally, the information about the sign of the impulse is encoded by incrementing the transmitted index by 128 when the sign of the glottal pulse is negative. The MSB in the 8-bit index thus represents the sign of the last glottal pulse. The inverse procedure is done at the decoder.

The phase control information is sent only in GC mode at 32 and 64 kb/s.

### 5.5.4 Pitch lag information

To improve speech quality under erroneous channel, a pitch lag estimate of the next frame is calculated at the encoder and transmitted as a side information for better excitation at concealed frame. By exploiting the 8.75-ms look-ahead signal used for the frame-end autocorrelation calculation, the pitch lag can be obtained without any additional delay. This tool is activated only at ACELP frame under operational modes of 24.4kbps.

The side information includes activation flag. For the frames classified as ONSET or VOICED under GC or VC mode, the activation flag is set to 1. For the other frames, the activation flag is set to 0.

In case the activation flag equals to 1, the pitch lag is encoded with 4 bits and transmitted on top of the activation flag. In case the activation flag equals to 0, only the activation flag is transmitted as side information.

To estimate a pitch lag at the look-ahead signal, this tool uses an extrapolated LSF parameter  and corresponding LP coefficients. For the LSF parameter extrapolation, the mean LSF vector is updated every frame and the extrapolated LSF  is calculated as follows.

 (1302)

where  is LSF vector of the last 3 frame, and  is the mean LSF vector. ,  depends on the previous coder type and the signal class. Decision rule for the constants used for LSF concealment applies to the decision of , . The extrapolated LSF  is converted into LSP parameter and extrapolated LP coefficients. The procedure is the same as the one under clean channel.

LP residual  is calculated for the 8.75-ms look-ahead signal with the extrapolated LP coefficients. LP residual  is used as target signal without perceptual weighting to estimate the pitch lag with low complexity. The pitch lag estimate is obtained by maximizing the following correlation.

 (1303)

where L' is the number of samples of the 8.75-ms look-ahead sub-frame, and  is the past excitation at the delay of *k*. The search range is limited to [,], where  is the pitch lag of the last sub-frame. Then the differential pitch lag from  is included to the side information with 4 bits.

### 5.5.5 Spectral envelope diffuser

A frame loss around speech onset sometimes causes too sharp peaks at LP spectrum, and sudden power increase in concealed signal. Spectral envelope diffuser mitigates the sudden power change and provides better recovery of concealed signal. The activation flag for spectral envelope diffuser is encoded with 1 bit and transmitted as a side information. This tool is active only at 9.6, 16.4, 24.4 kbps.

The activation is based on the function of merits depending on LSF improvement counter , quantized LSF parameter of the previous frame , the extrapolated LSF  obtained in the guided PLC for pitch lag at the previous frame, and a modified LSF parameter . The modified LSF parameter  is calculated as follows.

 (1304)

where  is the lowest number of *j* which satisfies the following equation.

 (1305)

 (1306)

where  is computed as follows.  is a threshold which equals to 1900 for 12.8 kHz internal sampling frequency, 2375 for 16 kHz internal sampling frequency.

After initialized with 0, the LSF improvement counter  is computed as follows:

(1307)

In case the following 4 equations are satisfied, the activation flag is set to 1, otherwise set to 0.  equals to 90 for 12.8 kHz internal sampling frequency, 112.5 for 16 kHz internal sampling frequency.  equals to 800 for 12.8 kHz internal sampling frequency, 1000 for 16 kHz internal sampling frequency.

(1308)

(1309)

(1310)

 (1311)

(1310)

 (1311)

The activation flag is updated based on algebraic codebook gain of the previous and the current frame. In case one of the following equations is satisfied, the activation flag is set to 0.

 (1312)

where  is the minimum value of algebraic codebook gains of current frame,  is the mean value of algebraic codebook gains of current frame, and  is the mean value of algebraic codebook gains of the previous frame.

The activation flag is further updated with the stability factor of LSF parameter. In case the stability factor is greater than 0.02, the activation flag is set to 0.

Finally the activation flag is encoded with 1 bit and transmitted as side information.

### 5.5.6 Tonality flag information

The flag  is set to one if the bit rate is one out of the set of {48 kbps, 96 kbps, 128 kbps}. For every frame for which  is one, two parameters of spectral flatness are computed as follows:

Let  be the sequence number of an arbitrary frame and denote the spectral flatness of the 'th frame. is defined as follows:

 (1312a)

where  is the geometric mean of the signal amplitudes,  is the arithmetic mean of the signal amplitudes,  is the MDCT coefficient at frequency point , and  is the number of the frequency points. The MDCT coefficients are either the original MDCT coefficients or the spectrum-shaped MDCT coefficients.

The original MDCT coefficients and the spectrum-shaped MDCT coefficients are used to compute two parameters of spectral flatness of this frame, denoted  and  respectively. For the frame with the mode of TCX20, the spectral flatness of the frame is computed by using the MDCT coefficients of the whole frame. For the frame with the mode of TCX10, the spectral flatness of the frame is computed by using the MDCT coefficients of the second sub-frame. In both cases, is the smaller value between the number of the frequency points of the MDCT coefficients and 200. If is smaller than a threshold , the flag of frame type is set to tonal type; otherwise, the flag of frame type is set to non-tonal type. If is smaller than another threshold , the flag of frame type is reset to tonal type. Then the obtained flag of frame type, together with the coded bit stream, is transmitted to the decoder side.

## 5.6 DTX/CNG operation

### 5.6.1 Overview

This subclause describes the discontinuous transmission (DTX) scheme and the comfort noise generation (CNG) algorithm. The DTX/CNG operation, which is activated on a command line, is used to reduce the transmission rate by simulating background noise during inactive signal periods. The regular DTX/CNG modes are supported for bit rates up to 24.4 kbps. For higher bit rates, the EVS codec supports a less aggressive DTX/CNG scheme that only switches to CNG for low input signal power.

The reduction of the transmission rate during inactive periods is achieved by coding the parameters referred to as comfort noise (CN) parameters. These parameters are used at the decoder to regenerate the background noise as well as possible, by respecting the spectral and temporal content of the background noise at the encoder. In the EVS Codec, the CNG algorithm reproduces high quality comfort noise by choosing between a linear prediction-domain based coding mode (LP-CNG) and a frequency-domain based coding mode (FD-CNG), according to the input characteristics. Each of the two coding modes utilizes a different set of CN parameters. In the LP-CNG mode, four CN parameters are analyzed and encoded: the low-band excitation energy, the low-band signal spectrum, the low-band excitation envelope and the high-band energy, where the high-band energy is only encoded for SWB/FB input. In the FD-CNG mode, the CN parameters consisting of global gain and spectral energies grouped in critical bands. Those parameters are encoded by a vector quantizer for transmission.

When the codec is operated with the DTX/CNG operation, the signal activity detector (SAD) is used to analyse the input signal to determine whether the signal comprises an active or inactive signal (see SAD decision in subclause 6.2). Based on its analysis, the SAD generates a SAD flag, , whose state indicates whether the signal is active (= 1) or merely a background noise (= 0). When = 1, the regular encoding and decoding process is performed, as in the default option. When = 0, DTX functions are run at the encoder that transmit either a silence insertion descriptor (SID) frame or a NO\_DATA frame. The SID frame contains the CN parameters, which are used to update the statistics of the background noise at the decoder, whereas the NO\_DATA frame is empty. The SID frame is always encoded using 48 bits regardless the actual CNG mode operating.

Further, hangover logic, as described in subclause 6.2, is used to enhance the quality of SID frames. The hangover logic in the SAD algorithm is such that the encoder waits for a certain number of frames before switching from the active signal to inactive signal. If the background noise contains transients that force the encoder to switch from inactive signal to active signal and then back to inactive signal in a very short time period, no hangover is used.

#### 5.6.1.1 SID update

The CN parameters are transmitted at a fixed or adaptive rate during inactive signal periods using a command line parameter. By default in the command line the transmission rate of CNG update is fixed to 8 frames. However, the CNG update rate can also be set to another fixed value or a variable rate by means of a command line parameter. The fixed rate is limited to between 3 and 100 frames. The adaptive rate, in general, is dependent on the background noise characteristics such as the current signal-to-noise ratio (SNR) and is limited to be between 8 and 50. Generally, at a high SNR, the SID frames are transmitted with a lower rate to achieve a significant reduction of average data rate at the cost of only minor quality degradation. On the other hand, at a low SNR, SID frames are transmitted with a higher frequency so that the comfort noise remains as natural as possible. Thus, increasing SNR implies decreasing SID frame frequency, whereas decreasing SNR implies increasing SID frame frequency.

To determine the adaptive SID transmission rate, the SNR is calculated based on the long-term energy of the active signal, , and background noise, . The DTX algorithm updates both long‑term values in each frame to take into account the possible evolutions of the level of the two respective signals. In the current frame, only one of these two energies is updated. If the current frame is classified as VOICED, the DTX module updates only the long-term energy of the active signal. Otherwise, it updates only the long-term energy of the background noise. The adaptive rate calculation is performed in every inactive signal frame after the preamble period. This period is characterized by at least 50 updates of both  and .

The update of the long-term energy of an active signal is performed as follows:

 (1313)

and the update of the long-term energy of an inactive signal as

 ()

where *Ef* = ||*snr*(*n*)|| is the energy of the denoised signal, *snr*(*n*), in the current frame. *α* represents a forgetting factor. Its value is based on the energy evolution. The value of *α* is set either to 0.99 for slow adaptation, or 0.90 for fast adaptation of the long-term energy level. In case of , fast adaptation is chosen if  in the current frame, otherwise slow adaptation is applied. In case of , the fast adaptation is chosen if  in the current frame, otherwise slow adaptation is applied.

Having estimated the long-term energies,  and , the SNR value in the logarithmic domain is calculated in every inactive signal frame as

 ()

The SID transmission rate, *rSID*, is finally adapted based on the current SNR value. The value of *rSID* is linearly varied between a minimum value, *rMIN*, that corresponds to a minimum SNR value, *SNRMIN*, and a maximum value, *rMAX*, that corresponds to a maximum SNR value, *SNRMAX*. That is

 ()

where *rMIN* ≤ *rSID* ≤ *rMAX*. The values *rMIN*, *SNRMIN*, *rMAX* and *SNRMAX* are selected as follows:

|  |
| --- |
| rMIN = 8  SNRMIN = 36 dB  rMAX = 50  SNRMAX = 51 dB. |

Thus, the adaptive rate is limited to between 8 and 50 frames and is updated in every inactive signal frame. If the number of consecutive NO\_DATA frames is equal to or greater than the current value of *rSID*, the next inactive signal frame is denoted as a SID frame. There is one exception to this rule.

1. SID frame sent due to detection of abrupt changes in the spectral characteristics of background noise, as described in Section 5.6.1.2.

After the rate *rSID* is determined, a variation of the long-term energy of the inactive signal is calculated and subjected to a fixed threshold. This is performed in every NO\_DATA frame after the preamble period. That is, if the following variation holds

 ()

the long-term energy of inactive speech is reset to , where  is the long-term energy of the inactive signal updated in every SID frame.

#### 5.6.1.2 Spectral tilt based SID transmission

Adaptive SID update rate that relies only on fluctuations in SNR as described in Section 5.6.1.1 may sometimes fail to detect significant changes in the background noise characteristics. In some cases, inactive frames that are perceptually different will have similar energy characteristics (typically encoded as gain values in the SID frames). Although background noise in a street (street noise) may have an energy distribution over time that is similar to that of background noise in a crowded space (babble noise), for example, these two types of noise will usually be perceived very differently by a listener. Spectral tilt is a good measure to capture such changes in the background noise characteristics.

It would be beneficial for a DTX/CNG scheme to track such perceptual changes in the background noise, apart from tracking the SNR as described in Section 5.6.1.1. Hence a scheme to detect a sudden change in spectral tilt of the background noise and trigger a new SID frame indicating the parameters of the new background noise is employed in the encoder.

To ensure that there is no affect from an active talk spurt to the computation of spectral tilt of the background noise, this computation is performed in every inactive frame after 5 consecutive inactive frames that immediately follow an active talk spurt.

Linear prediction coefficients (LPCs) are derived using performing Linear prediction analysis (Section 5.1.5.1 – Section 5.1.5.3) on the current input frame with background noise and no active speech. LPCs are then converted to reflection coefficients (RCs) as follows using a backwards Levinson Durbin recursion. For a given Nth order LPC vector, the Nth reflection coefficient value is derived using the formula , it is then possible to calculate the lower order LPC vectors using the following recursion

 ()

which yields the reflection coefficient vector . The spectral tilt of background noise is indicated by the first reflection coefficient. A smoothened running average of the spectral tilt of background noise in *K*th inactive frame is computed using a first order IIR filter as follows.

 ()

The running average differences from frame to frame are accumulated in during each during each successive inactive frame *K* as follows:

 ()

Absolute value of this delta-sum parameter is compared against a set threshold of 0.2. If this threshold is exceeded during an inactive frame indicating a change in spectral tilt characteristics of the background noise, and if the number of consecutive NO\_DATA frames is equal to or greater than 8, a new SID frame is transmitted regardless of the current value of the adaptive SID rate parameter *rSID*. At this point, parameters and are also reset to zero to start a fresh computation of spectral tilt of the background noise that follows this SID frame.

#### 5.6.1.3 CNG selector

The CNG selector chooses one of the two CNG modes (FD\_CNG or LP\_CNG) for generating comfort noise. In case AMR-WB IO mode is used, LP-CNG is always selected. Otherwise, the decision is based on the energy ratio between a high and a low frequency range of the background noise signal and the bandwidth of the signal.

Noise energy estimates for the low frequency range up to 1270 Hz and for the two highest critical bands are estimated by , ()

, ()

except for narrowband signals, where the lowest band is ignored and the highest bands are lower:

, ()

. ()

 is the background noise energy per critical band as described in subclause 5.1.11.1. Both values and are used to calculate the spectral tilt of the background noise energies and update the memory for:

 ()

Depending on the previous CNG mode, input signal bandwidth (input\_bwidth) as detected by the bandwidth detection module (subclause 5.1.6) and the CNG mode is changed if the current frame is active and at least the past 20 frames were active and one of the following cases applies:

if (cng\_mode == LP\_CNG &&

(( input\_bwidth == NB &&  > 9.f) ||

( input\_bwidth > NB &&  > 45.f)))

{

cng\_mode = FD\_CNG;

}

else

if ( cng\_mode == FD\_CNG &&

(( input\_bwidth == NB &&  < 2.f) ||

( input\_bwidth > NB &&  < 10.f)))

{

cng\_mode = LP\_CNG;

}

### 5.6.2 Encoding for LP-CNG

This section describes the operation in LP-CNG. Similar to the default operation, the LP-CNG also operates on the split-band basis. In WB/NB operation, only the LP parameters for low-band signal are analyzed and encoded. In SWB/FB operation, besides the LP analysis for the low-band signal, the high-band signal is analyzed and encoded separately as a kind of bandwidth extension. The LP parameters for low-band analysis include: the low-band excitation energy, the low-band signal spectrum and the low-band excitation envelope. The high-band analysis only involves one parameter which is the high-band energy. The 1 CNG type bit (see subclause 7.2) is set to “0” for LP-CNG and transmitted in each SID frame.

#### 5.6.2.1 LP-CNG CN parameters estimation

The CN parameters to be encoded into a LP-CNG SID frame are calculated over a certain period, which is called the CN averaging period. These parameters give information about the level and the spectrum of the background noise. The CN averaging period, *NCN*, is equal to the number of consecutive frames including the current SID frame and its preceding NO\_DATA frames, upper-limited by the value of 8 consecutive frames. It is a variable value depending on the current SID transmission rate. In particular, the first SID frame immediately after an active signal burst always uses the value *NCN* = 1. The LP-CNG can generate two different types of SID frame – the WB SID frame, containing low-band (WB) only CN parameters, and the SWB SID frame, containing both low-band and high-band (SHB) CN parameters. One bit is encoded into the LP-CNG SID frame to indicate the bandwidth type of the SID frame, where “0” indicates WB SID and “1” indicates SWB SID. Only WB SID frames are transmitted when operating in NB/WB mode. In SWB/FB operation, the SWB SID frames are not always transmitted but WB SID frames can also be transmitted between two adjacent SWB SID updates. This means the high-band CN parameters are not updated at the decoder with the same rate the low-band CN parameters will be updated. Details in SWB/FB operation will be described in subclause 5.6.2.1.8. The bit allocation for the CN parameters in the respect WB and SWB SID frames are described in subclause 7.2.

##### 5.6.2.1.1 LP-CNG Hangover analysis period determination

To enable high quality comfort noise synthesis on the receiving side, the encoder sends a three bit counter value to the decoder. The transmitted value is derived from the counter, as:

 ()

where is counter of the consecutive frames without any primary SAD active decisions where the DTX SAD flag , as described in subclause 5.1.12.8, is set to 1. These hangover frames without primary SAD active decisions are deemed to be relevant for comfort noise analysis. For DTX operation, the  counter is incremented in actively encoded speech segments whenever the primary SAD flag is set to 0 and the DTX SAD flag is set to 1 and it is set to zero whenever this is not the case.

##### 5.6.2.1.2 LP-CNG filter parameters evaluation for low-band signal

In the DTX/ LP-CNG operation, the LP filter coefficients are quantized in the LSF domain, which is the same as in the default option. However, in the case of DTX/CNG operation, the LP filter coefficients are not interpolated within the frame. From the LP analysis, which is described in subclause 5.1.9, only the end-frame LSP vector, , is used for quantization purposes in SID frames.

The end-frame LSP vector is not quantized directly. Instead, an averaged end-frame LSP vector is calculated over the CN averaging period which is then converted to an LSF vector and quantized. Not all end-frame LSP vectors in the CN averaging period are used for averaging, but two outlier LSP vectors are removed. The two outliers are found over the CN averaging period, as the two LSP vectors representing the lowest spectral entropy. This is to mitigate the possible corruption to the averaged LSP vector from interfering background frames, assuming that interfering background frames are usually more structural in their spectrum (leading to lower spectral entropy) than normal background noise frames. A parameter  which can reflect such a spectral entropy is thus calculated for each end-frame LSP vector over the CN averaging period. Each end-frame LSP vector is first converted to its LSF representation and  is calculated as

 ()

where  equals 16 which is the order of the LP filter,  denotes the bandwidth of the partition if the signal bandwidth is divided by M equally spaced LSF coefficients, that is, , where  is the bandwidth of the signal,  equals 6400 for 12.8kHz sampling rate core and 8000 for 16kHz sampling rate core,  denotes the length of the  partition divided by LSF coefficients of the  LSP vector over the signal bandwidth, that is

 ()

where  is the  LSF coefficient of the  LSP vector,  is either 6400 or 8000 depending on the sampling rate of the core. A more structured spectrum will result in a  having higher value representing lower spectral entropy. So the two LSP vectors resulting in the two maximum  over the CN averaging period is found as the two outliers. The averaged LSP vector is then calculated as

 ()

where  is the length of CN averaging period, , denotes the index of the two LSP outliers. The outlier-removed LSP vector average  is considered the best representation of the short-term spectral envelope of the background noise. It is converted to the LSF representation, quantized (see subclause 5.6.2.1.3) and transmitted in the SID frame.

##### 5.6.2.1.3 LP-CNG CNG-LSF quantization for low-band signal

The quantization of the LSF vector follows the procedures used for the LSF vector within the ACELP block. They are described in subclause 5.2.2.1. The quantization is done with a two stage quantizer. The first stage consists of a non-predictive, non-structured, optimized VQ codebook. The second stage consists of a multiple scale lattice vector quantizer whose structure and search procedure are detailed in subclause 5.2.2.1.4. The number of lattice structures from the second stage corresponds to the number of codevectors from the first stage such that if a particular codevector is selected in the first stage, its corresponding lattice structure is used in the second stage. A lattice multiple scale lattice structure corresponds to a set of 6 numbers specifying the number of leader classes in each of the 6 lattice truncations and 6 numbers specifying the scale values for the same truncations. There are three lattice truncations that define the codebook for the first 8-dimensional LSF subvector and three lattice truncations defining the codebook for the second 8-dimensional LSF subvector. In addition, a 16-dimensional vector defines for each multiple scale lattice structure a normalization values vector, .

The codebook from the first stage uses 4 bits, and each lattice structure from the second stage is defined for 25 bits. Consequently a total of 29 bits that are used for the quantization of the LSF vector. With the above described structure, the table ROM used for storing the CNG LSF codebook data covering 29 bits has 1.408kBytes.

The search in the first stage codebook is done taking into account the value of the last component of the 16 dimensional LSF vector. Based on this value only part of the first stage codebook is searched. If the last LSF vector component is larger than 6350 then the search is done only for the first 6 codevectors of the first stage and the LSF vector corresponds to internal sampling frequency of 16kHz, otherwise the search is performed within the last 10 codevectors of the first stage that correspond to the internal sampling rate of 12.8kHz. At the second stage, prior to quantization with the lattice structure, based on the selected first stage codevector some components of the LSF vector are permuted like specified by the following table:

Table 151: LSF vector component permutation

|  |  |
| --- | --- |
| First stage codevector index | Permutations |
| 0 | (6,11), (7,15) |
| 1 | (6,15) |
| 2 | (5,8), (7,15) |
| 3 | (7,10) |
| 7 | (0,9), (7,10) |
| 9 | (7,15) |
| 12 | (6,10), (7,11) |
| 13 | (6,10), (7,12) |
| 14 | (6,10), (7,12) |
| 15 | (6,10), (7,12) |

A permutation defined as (6,11) signifies that the 6th component, numbered starting from 0, is replaced by the 11th one and reciprocally. The permutations are performed between the two groups or subvectors, i.e the first index in the permutation is from the first half of the codevector and the second one from the second half or subvector. The permutations are performed only when one of the first stage codevectors whose index is mentioned in the previous table is obtained at the first stage. The resulting vector is component wise multiplied with the inverse of the corresponding vector and quantized with the corresponding multiple scale lattice structure. After quantization the components of the obtained second stage multiple scale lattice codevector are permuted back, and added to the first stage codevector in order to obtain the quantized LSF vector. 1 bit indicating the core sampling rate is transmitted in each SID frame. This bit signals the decoder the sampling domain on which the quantized LSF vector is. The bit is set to “0” for 12.8 kHz sampling rate and “1’ for 16 kHz sampling rate.

##### 5.6.2.1.4 LP-CNG synthesis filter computation for local CNG synthesis

A smoothed LSP vector is used in every inactive frame to obtain the LP synthesis filter .The quantized LSF vector is converted back to the LSP domain. The smoothed LSP vector is updated by the last quantized LSP vector by means of an AR low-pass filter in each inactive frame except the first SID frame after an active burst. That is

 ()

where , denote respectively the smoothed LSP vector at the current and the previous frame,  denotes the last quantized LSP vector, = 0.9 is a smoothing factor. An additional constraint is applied to the inactive frames after the first SID frame of an inactive segment and before the second SID frame that the update to the smoothed LSP vector described above is suspended if the last SID excitation energy is an outlier and sufficient hangover frames are contained in the last active burst, that is, if  and , where  denotes the quantized excitation energy in the last SID frame, as calculated in equation (1339),  denotes the number of entries in used for  calculation as described in subclause 6.7.2.1.2 and  is the smoothed quantized excitation energy, further described in subclause 5.6.2.1.6. For the first SID frame after an active burst, the smoothed LSP vector is updated depending on whether the frame is an outlier in either energy or spectrum and whether there are past CN parameters to analyze in the CNG analysis buffer as described in subclause 6.7.2.1.2. If the step update flag  is set to 1 or there were no past CN-parameters to analyse in the CNG analysis buffer, the smoothed LSP vector is initialized to the quantized LSP vector of the current SID frame. Otherwise, if step update is not allowed and there are past CN parameters to analyze in the CNG analysis buffer, the overall and the maximum individual spectral distortion between the quantized LSP vector of the current SID frame and the average LSP vector, , calculated over hangover frames in subclause 6.7.2.1.2 are calculated.

 ()

 ()

where  is the overall spectral distance,  is the maximum spectral distance,  is the average LSP vector calculated over hangover frames,  is the quantized LSP vector of the current SID frame. If  and  are deviating to each other, that is, if  or , the quantized LSP vector of the current SID frame , is considered an outlier and the smoothed LSP vector is initialized to the average LSP vector calculated over hangover . Otherwise, the smoothed LSP vector is initialized by

 ()

The smoothed LSP vector  is initially set to the quantized end-frame LSP vector from the previous frame, , when the first SID frame is processed at the encoder. The step update flag is set in each inactive frame by measuring the energy step between the instant energy and the long-term energy. For the first inactive frame after an active signal period, the flag  is additionally set if there are past CN-parameters and the most recent energy value in  is more than four times larger than the smoothed quantized excitation energy . Finally, the smoothed LSP vector is converted to LP coefficients to obtain a smoothed LP synthesis filter, , which is used in the local CNG synthesis.

##### 5.6.2.1.5 LP-CNG energy calculation and quantization

The excitation energy in the current frame is computed for each inactive frame according to the following equation:

 ()

where  is the LP residual signal, calculated by filtering the pre-emphasized inactive input signal, , through the filter *Â*(*z*), =256 or 320 depending on the sampling rate of the core. Then a weighted average energy is computed over the whole CN averaging period by

 ()

where the weights are defined as  = [0.2, 0.16, 0.128, 0.1024, 0.08192, 0.065536, 0.0524288, 0.01048576], and  is an energy offset value which is set to 0 for input bandwidth = NB, to 1.5 for input bandwidth greater than WB, and for signals of bandwidth = WB, the energy offset value is chosen from an energy attenuation table depending on the latest bitrate used for actively encoded frames  as defined by Table 151a. The energy offset is only updated in the first SID frame after an active signal period if two criteria are both fulfilled. The first criterion is satisfied if AMR-WB IO mode is used or the bandwidth=WB. The second criterion is met if the number of consecutive active frames in the latest active signal segment was at least number of frames or if the current SID is the very first encoded SID frame. The superscript [*n*] denotes a particular frame, e.g., [0] is the current frame.

Table 151a: Energy offset selection for LP-CNG

|  |  |
| --- | --- |
| Latest active bitrate [kbps] |  |
|  | 1.7938412 |
|  | 1.3952098 |
|  | 1.0962363 |
|  | 0.9965784 |
|  | 0.9965784 |

The weighted average energy is then quantized using a 7-bit arithmetic quantizer. The integer quantization index in the current SID frame is found using the relation

 (1336)

where Δ = 5.25 is the quantization step. The quantization index is limited to [0, 127]. The quantization index is further limited not to increase by more than one from the value of the previous frame if the previous frame was also an inactive frame. An exception is that if the step update flag  is set to 1, then the quantization index is allowed to increase more than one from the value of the previous frame using the relation

 ()

where  denotes the final quantization index transmitted in each SID frame,  denotes the quantization index transmitted in the previous SID frame, is the quantization index calculated in equation (1336). The quantized value of energy is used further in the local CNG synthesis and is found by

 ()

which is converted to the linear domain by

 (1339)

##### 5.6.2.1.6 LP-CNG energy smoothing for local CNG synthesis

The quantized excitation energy, , calculated in equation (1339) is not used directly in the local CNG synthesis. Instead, a smoothed quantized excitation energy,, is computed. The smoothed quantized excitation energy is updated in every inactive frame in a general form of

 ()

where superscript [-1] demotes the value from the previous frame,  is the quantized energy in the SID frame, calculated in equation (1339),  is the smoothing factor controlling the update rate. Variable update rates are utilized. For the first inactive frame after an active signal period, if there is no step update flag  set to 1 in the latest two SID frames,= 0.8 if the number of preceding hangover frames is less than 3 or , otherwise, = 0.95. Otherwise, if step update flag  is set to 1 in either of the latest two SID frames, = 0, i.e.  is set directly to . For consequent frames, if the step update flag  at the latest SID frame is not set to 1, = 0.8. Otherwise, = 0, i.e.  is set directly to . For the first inactive frame after an active signal period, before the smoothed quantized excitation energy  is updated by , the value of  from the previous frame, that is , is initialized to  which is the age weighted average excitation energy of the DTX hangover frames calculated in subclause 6.7.2.1.2, if step update flag  is not set to 1 and there are past CN parameters to analyse in the CNG analysis buffers. is initially set equal to .

##### 5.6.2.1.7 LP-CNG LF-BOOST determination and quantization

While the quantized LSF spectrum generally estimates well the spectrum of most background noises, it is found less sufficient for noises which have strong low frequency component for example the car noise. To compensate the missing low frequency component, the spectral envelope in the low frequency portion of the LP residual signal is quantized and transmitted in the SID frame. Note that this quantized residual spectral envelope is only transmitted in WB SID frame

The LP residual signal  calculated in subclause 5.6.2.1.5 is first attenuated by multiplying an attenuation factor  for all input bandwidth except NB. The attenuation factor is calculated as

 ()

where  if the bandwidth is not WB or the latest bitrate used for actively encoded frames  is larger than 16.4 kbps. Otherwise  is determined from a hangover attenuation floor table as defined in table 35b. The attenuation factor  is finally lower limited to. Then a FFT is used to obtain the frequency representation of the LP residual signal and a spectral envelope which is the energies of the first 20 FFT bins in the low frequency portion of the frequency representation (excluding the DC bin) is calculated as

 (1342)

where andare, respectively, the real and the imaginary parts of the-th frequency bin as outputted by the FFT, = 256 is the size of FFT analysis. This low frequency spectral envelope of the LP residual is not quantized directly. Instead, an averaged spectral envelope is calculated over the CN averaging period. The averaging is similar to the process in subclause 5.6.2.1.2 that the spectral envelopes of the two outliers identified in subclause 5.6.2.1.2 are removed from the averaging. The averaged low frequency spectral envelope is calculated as

 ()

where  is the length of CN averaging period,  denotes the low frequency envelope of the -th frame in the CN averaging period, , denotes the index of the two outliers. To encode this averaged low frequency spectral envelope, the spectral details of the averaged low frequency spectral envelope is extracted and used for actual quantization. The spectral details,, is obtained by subtracting an envelope floor which is equal to two times of the quantized average excitation energy from the averaged low frequency spectral envelope, that is

 ()

whereis bounded to non-negative value.  is then converted to log domain

 ()

where  is the energy offset value calculated in subclause 5.6.2.1.5 and are bounded to non-negative value. A distance vector is calculated as

 ()

where  is the quantized total excitation energy calculated as

 ()

where = 256 is the length of the excitation. The distance vector  is quantized by a vector quantization. The codeword having the minimum prediction error is found by a direct search in the codebook and the index of the codeword is transmitted in the WB SID frame. The quantized low frequency envelope is recovered and used in the local CNG synthesis.

##### 5.6.2.1.8 LP-CNG high band analysis and quantization

To enable high perceptual quality in the inactive portions of speech on the decoder side, during SWB mode DTX/LP-CNG operation of the codec, the high band noise signal (6.4 - 14.4 kHz or 8 - 16 kHz depending on the core sampling rate) is analyzed and quantized. However, this is being done without transmitting any extra parameters from the encoder to decoder to model the high-band spectral characteristics of the inactive frames. Instead, the high band spectrum of the comfort noise is modelled purely in the decoder side. Only the energy of high band signal is quantized and transmitted in the SID frame.

The average energy of the high band signal is first calculated

 ()

where  is the high band signal, = 320 is the length of high band signal. The log average energy of high band signal is calculated and to which an attenuation of 6.5 dB is applied

 ()

The attenuated log energy  is smoothed by an AR filtering as

 ()

where  is the smoothed high band log average energy,  denotes the smoothed log average energy in the previous frame. The average energy of the low band signal is also calculated

 ()

where  is the synthesized low band signal as described in subclause 5.6.2.2,  is the length of the synthesized low band signal. The log average energy of the low band signal is calculated

 ()

The log average energy of the low band signal  is also smoothed by an AR filtering.

 ()

where  is the smoothed low band log average energy,  denotes the smoothed log average energy in the previous frame. Step update to the smoothed low band and high band log average energy is allowed. If the low band log average energy  of the current frame is deviating from the smoothed low band log average energy of the previous frame  by more than 12dB, a step update flag  is set to 1 indicating the permission of step update, otherwise is set to 0. If  is set to 1, the smoothed low band log average energy and the smoothed high band log average energy are respectively set to the low band log average energy  and the high band log average energy . The high band parameter, i.e. the energy of the high band signal is not quantized and transmitted in every SID frame. Instead, SWB SID frame which contains both the low band and high band parameters is only transmitted when the energy relationship (the energy ratio) between the low band and high band signals at the current frame is deviating from that relationship at previous SWB SID frame by more than 3dB. This can be described as, when a SID frame is about to be transmitted, if , then the high band parameter is transmitted in the SID frame. Besides, following conditions can also trigger the transmission of SWB SID frame, including: the first SID frame immediately after active frames, the SID frame which is within high band analysis initialization period, the SID frame before which there is no active and no SWB SID frame in the 100 preceding frames when operating in SWB mode or above, the SID frame where there is bandwidth switching between WB and SWB. If SWB SID transmission is not triggerd at an instance of SID frame update, the WB SID frame will be transmitted instead.

In each SWB SID frame, the smoothed high band log average energy  is quantized and transmitted. The  is first converted to  domain as

 ()

Then a 4-bit arithmetic quantizer is used for quantizing . The integer quantization index is found by

 ()

where = 0.9 is the quantization step. The quantization index  is bounded to [0, 15].

#### 5.6.2.2 LP-CNG local CNG synthesis

The local CNG synthesis is performed at the encoder for low-band signal in order to update the filters, the adaptive codebook memories, and to guide the high-band analysis and the DTX hangover control (see subclause 5.1.12.8). The local CNG is performed by filtering a scaled excitation signal through a smoothed LP synthesis filter. The scaled excitation,, is a combination of a random excitation and an excitation representing the low frequency spectral details of the excitation signal. For the generation of , see subclause 6.7.2.1.5. For the computation of the smoothed LP synthesis filter, see subclause 5.6.2.1.4.

#### 5.6.2.3 LP-CNG CNG Memory update

When an inactive signal frame is encoded, the following updates are carried out:

– MA memory of the ISF quantizer is set to zero;

– AR memory of the ISF quantizer is set to its mean values (UC mode, WB case);

– synthesis excitation spectrum tilt is set to zero;

– weighting filter denominator memory is set to zero;

– gain of pitch clipping memory is set to initial values;

– open-loop pitch estimator parameters are set to zero;

– per-bin NR last critical band is set to zero (the whole spectrum subtraction);

– noise enhancer memory is set to zero;

– phase dispersion memory is set to zero;

– previous pitch gains are all set to zero;

– previous codebook gain is set to zero;

– voicing factors used by bandwidth extension are all set to 1;

– active frame counter is set to zero;

– bass post-filter is tuned off;

– floating point pitch for each subframe is set to the subframe length;

– class of last received good frame for FEC is set to UNVOICED\_CLAS;

– synthesis filter memories are updated.

### 5.6.3 Encoding for FD-CNG

To be able to produce an artificial noise resembling the actual input background noise in terms of spectro-temporal characteristics, the FD-CNG makes use of a noise estimation algorithm to track the energy of the background noise present at the encoder input. The noise estimates are then transmitted as parameters in the form of SID frames to update the amplitude of the random sequences generated in each frequency band at the decoder side during inactive phases. Note, however, that the noise estimation is carried out continuously on every frame, i.e., regardless of the speech activity. Therefore, it can deliver some meaningful information about the noise spectrum at any time, in particular at the very beginning of a speech pause.

The FD-CNG noise estimator relies on a hybrid spectral analysis approach. Low frequencies corresponding to the core bandwidth are covered by a high-resolution FFT analysis, whereas the remaining higher frequencies are captured by the CLDFB which exhibits a significantly lower spectral resolution of 400Hz.

The size of an SID frame is however very limited in practice. To reduce the number of parameters describing the background noise, the input energies are averaged among groups of spectral bands called partitions in the sequel.

#### 5.6.3.1 Spectral partition energies

The partition energies are computed separately for the FFT and CLDFB bands. The energies corresponding to the FFT partitions and the  energies corresponding to the CLDFB partitions are then concatenated into a single array of size  which will serve as input to the noise estimator described in subclause 5.6.3.2.

##### 5.6.3.1.1 Computation of the FFT partition energies

Partition energies for the frequencies covering the core bandwidth are obtained as

, (1356)

where  and  are the average energies in critical bandfor the first and second analysis windows, respectively, as explained in subclause 5.1.5.2. The number of FFT partitions  depends on the sampling rate of the input signal, as show in Table 133. The de-emphasis spectral weights are used to compensate for the high-pass filter described in subclause 5.1.4 and are defined as

 (1357)

##### 5.6.3.1.2 Computation of the CLDFB partition energies

The partition energies for frequencies above the core bandwidth are computed as

, (1358)

where and  are the indices of the first and last CLDFB bands in the *i*-th partition, respectively,  is the total energy of the *j*-th CLDFB band (see subclause 5.1.2.2), and is a scaling factor computed in subclause 5.1.6.1. The constant 16 refers to the number of time slots in the CLDFB. The number of CLDFB partitionsdepends on the configuration used, as described in the next subclause.

##### 5.6.3.1.3 FD-CNG configurations

The following table lists the number of partitions and their upper boundaries for the different FD-CNG configurations at the encoder, as a function of the input sampling rate.

Table 152: Configurations of the FD-CNG noise estimation at the encoder

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| (kHz) |  |  | (Hz) | (Hz) |
| 8 | 17 | 0 | 100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700 |  |
| 16 | 20 | 1 | 100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6350 | 8000 |
| 32/48 | 20 | 4 | 100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6350 | 8000, 10000, 12000, 16000 |

For each partition,corresponds to the frequency of the last band in the *i*-th partition. The indicesand  of the first and last bands in each spectral partition can be derived as a function of the common processing’s sampling rate 12.8 kHz and FFT size 256 (see subclause 5.1):

, ()

, ()

whereis the frequency of the first band in the first spectral partition. Hence the FD-CNG generates some comfort noise above 50Hz only.

#### 5.6.3.2 FD-CNG noise estimation

The FD-CNG relies on a noise estimator to track the energy of the background noise present in the input spectrum. This is mostly based on the minimum statistics algorithm [R. Martin, Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics, 2001].

However, to reduce the dynamic range of the input energies  and hence facilitate the fixed-point implementation of the noise estimation algorithm, a non-linear transform is applied before noise estimation (see subclause 5.6.3.2.1). The inverse transform is then used on the resulting noise estimates to recover the original dynamic range (see subclause 5.6.3.2.3). The resulting noise estimates are used in subclause 5.6.3.5 to encode the SID frames.

##### 5.6.3.2.1 Dynamic range compression for the input energies

The input energies are processed by a non-linear function and quantized with 9-bit resolution as follows:

. ()

Background of using log2 is that the (int)log2 can usually be calculated very quickly (in one cycle) on fixed-point processors using the “norm” function which determines the numbers of leading zeros in a fixed point number.

Background for adding a constant 1 inside the log2 function is to ensure that the converted energiesremain positive. This is especially important as the noise estimator rely on a statistical model of the noise energy. Performing noise estimation on negative values would strongly violate the model and can result in unexpected behaviour.

##### 5.6.3.2.2 Noise tracking

The input energy corresponds to an instantaneous power for the *i*-th partition, referred to as periodogram in the sequel. The minimum statistics algorithm relies on an optimally smoothed periodogramwhich can be considered as an estimate of the input power spectral density. The algorithm derives therefore an estimate of the noise power spectral density which we denote in the following as. As described in the sequel, some additional smoothing ofis applied, yielding the smoothed noise estimate  introduced in subclause 5.6.3.2.2.5.

5.6.3.2.2.1 Initialization phase

To correctly initialize the noise estimation algorithm, an initialization phase is used as long as the input energy  of the first partition grows. Note that the initialization phase is also triggered when a reset of the noise estimation algorithm is judged necessary, as described in subclause 5.6.3.4.

The following applies for each partition  during the initialization phase:

. ()

Moreover, the minimum statistics algorithm includes a bias compensation mechanism exploiting statistical moments  and of first and second orders, respectively. During the initialization phase, we have  
 and.

It is also necessary to initialize several summed quantities. The total noise energy over the FFT and CLDFB partitions are computed during the initialization phase as

 ()

and

, ()

respectively. Note that the quantity corresponds to the size of the *i*-th partition. The total input energies  and  for the FFT and CLDFB partitions are initialized to and, respectively, and the total smoothed input energies and  are initialized with and, respectively.

In subclause 5.6.3.2.2.4, some auxiliary arrays, , and  are also required to track minima in each spectral partition. They are all filled with the largest possible platform value during the initialization phase.

5.6.3.2.2.2 Optimal smoothing of the input energies

As mentioned earlier, the power spectral density estimate  is computed iteratively as a smoothed version of the input energy, i.e.,

 (1365)

whereis a time-varying optimal smoothing parameter. It is computed separately for the FFT and CLDFB:

 (1366)

where

(1367)



 (1368)

are some correction factors,

 (1369)

 (1370)

impose a lower limit on the optimal smoothing parameter, and

, (1371)

, (1372)

, (1373)

 (1374)

denote some summed quantities.

5.6.3.2.2.3 Bias compensation

The minimum statistics algorithm essentially consists in tracking the minima offor each partition *i* over time. However, this method delivers some biased estimates and necessitates therefore the computation of a bias compensation factor which is dependent on the variance of.

For each spectral partition, we first estimate the first-order moment of  as

, (1375)

where is a smoothing parameter. The variance of is then derived as follows:

. (1376)

As shown in subclause 5.6.3.2.2.4, the minimum tracking uses in each partition *i* a window of sub-windows of length each. Minima are in fact computed over the entire buffer of size past frames, but also over the last sub-window. The bias compensation factorfor the total window length, and  for a sub-window are given for each partition as

, (1377)

, (1378)

with

, (1379)

, (1380)

. (1381)

For the sake of robustness, the bias compensation factors are furthermore increased proportionally to the mean of  and  among the spectral partitions. The correction factor is obtained for the FFT and CLDFB partitions as

, (1382)

, (1383)

with

, (1384)

 (1385)

5.6.3.2.2.4 Minimum tracking

For the sake of simplicity, we provide a description of the minimum tracking algorithm for the FFT partitions only. The CLDFB partitions can be treated in the same way.

The bias compensation factor and the correction factor computed in subclause 5.6.3.2.2.3 are used to obtain a more accurate estimate of the background noise energy in each partition. They are re-computed after each frame and tracking of the minimum  is carried out in each FFT partitionas . When a new minimumis found, a flag is set to 1 and is updated as. Otherwiseis set to zero.

Note thatis set to the maximum possible platform value after processing the last frame of each sub-window, i.e., every frames.  and refer therefore to a minimum within the current sub-window for the partition *i*. In the last frame of the current sub-window, the current minimum is stored into a buffer collecting the minima found in the last  sub-windows. The buffer is used at the end of each sub-window to determine the overall minimum among all sub-windows.

For a frame in between the first and last frames of the current sub-window (i.e., frames 2 to  of the sub-window), the overall minimumis updated as, and a flag  is set to 1 if a new minimum was found, i.e., if. The flag is set to 0 after processing the last frame of each sub-window.

To improve tracking of a time-varying noise, a local minimum among the current sub-window can replace the overall minimum  in the last frame of each sub-window provided that it yields only a moderate increase of, and if the local minimum was not found in the first or last frame of the sub-window, i.e., if and. In this case, replaces also all values in the buffer. The search rangefor the local minima (and hence the tolerated amount of increase compared to the current overall minimum) lies between 1.1 and 2 and increases as  decreases:

. (1386)

The noise estimate is updated to  after each frame, except for the first frame in each sub-window.

Furthermore, when the smoothed energy of the first spectral partition exceeds the instantaneous energy by a factor of more than 50 (i.e., ) for at least two frames in a row, a sudden noise offset is assumed and the noise tracking is modified for all partitionsas:

, (1387)

, (1388)

, (1389)

and

 , (1390)

, (1391)

. (1392)

5.6.3.2.2.5 Smoothing of the noise estimates

The main outputs of the noise tracker are the noise estimates. To obtain smoother transitions in the comfort noise, a first-order recursive filter is applied, i.e.

. (1393)

Furthermore, the input energyis averaged over the last 5 frames. This is used to apply an upper limit on in each spectral partition.

##### 5.6.3.2.3 Dynamic range expansion for the estimated noise energies

The estimated noise energies are processed by a non-linear function to compensate for the dynamic range compression applied in subclause 5.6.3.2.1:

 (1394)

#### 5.6.3.3 Adjusting the first SID frame in FD-CNG

Before encoding the first SID frame of a CNG phase (i.e., an SID frame preceded by an active frame), an upper limit is applied to the noise estimates to minimize the risk of generating some noise bursts at the beginning of a CNG phase. To this end, the noise estimate obtained during the previous inactive frame (i.e., one frame before the last active phase) is used in combination with the input energy of the current frame as follows:

, (1395)

where  refers to the partition index, , and *num\_active\_frames* corresponds to the length of the last active phase.

#### 5.6.3.4 FD-CNG resetting mechanism

To signal the need of resetting FD-CNG, a flag is computed based on a detection of fast increasing noise energy.

If the current total noise energy  as calculated in subclause 5.1.11.1 is bigger than the one of the last frame, up to four difference values of the total noise energy of the previous frames are summed up, in case the noise energy increased in these last frames consecutively.

If the encoder is out of initialisation phase, four or more frames with increasing  were detected and the sum of the last four of them was bigger than 5, the reset flag is set to 1. Besides that, in case the signal’s input bandwidth of the current frame is larger than the previous one, the reset flag is also set to 1.

If none of this is the case but the flag was set before, it takes nine more frames before the flag is set to zero.

In case the flag is set to one, a reinitialization of the minimum statistics routine is triggered (subclause 5.6.3.2.2.1), and selection of SID or NO\_DATA frame is forbidden.

#### 5.6.3.5 Encoding SID frames in FD-CNG

The CN parameters to be encoded into a FD-CNG SID frame are the noise estimates, with, and  depends on the bandwidth and the bitrate as given in the table below.

Table 153: Number of CN parameters encoded in a FD-CNG SID frame

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bandwidth | NB | WB | | SWB |
| Bit-rates [kbps] |  |  |  |  |
|  | 17 | 20 | 21 | 24 |

The noise estimates  are first converted to dB

. ()

The noise estimates in dB are then normalized using

. ()

The normalized noise estimates in dB  are then quantized using a Multi-Stage Vector Quantizer (MSVQ). The MSVQ has 6 stages, with 7 bits in the first stage and 6 bits in the other stages (total of 37 bits). A M-best search algorithm is used, with M=24 is the number of survivors in a stage that will be searched in the next stage. Note that a single set of codebooks is used for all configurations. The vectors in the codebook have a length of 24, and they are simply truncated if  is less than 24.

The MSVQ decoder output is given by

, ()

where  are the indices encoded in the bitstream and  is the -th coefficient of the -th vector in the codebook of stage .

A global gain is then computed

, ()

with  is a scale which depends on the bandwidth and the bitrate as described in the table below.

Table 154: FD-CNG SID global gain scale

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Bandwidth | NB | | | | WB | | | | | | SWB | | |
| Bit-rates [kbps] | <=7.2 | 8 | 9.6 | 13.2 | <=7.2 | 8 | 9.6 | 13.2 | 16.4 | 24.4 | 13.2 | 16.4 | 24.4 |
| [dB] | -5.5 | -5 | -4 | -3 | -5.5 | -5 | -1.55 | -3 | -0.6 | -0.2 | -3 | -0.8 | -0.25 |

The global gain is then quantized on 7 bits using

, ()

producing the quantized global gain

. ()

The quantized noise estimates are then given by

. ()

Finally the last band parameter is adjusted in case the encoded last band size is different from the decoded last band size



#### 5.6.3.6 FD-CNG local CNG synthesis

Noise estimates encoded in the SID frame are then used to locally generate CNG, in order to update the encoder memories. The following table lists the number of partitions used for generating CNG in the FD-CNG encoder, and their upper boundaries, as a function of bandwidths and bit-rates.

Table 155: Configurations of the FD-CNG local CNG synthesis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Bit-rates  (kbps) |  |  | (Hz) | (Hz) |
| NB |  | 17 | 0 | 100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3975 |  |
| WB |  | 20 | 0 | 100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375 |  |
|  | 20 | 1 | 100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375 | 8000 |
|  | 21 | 0 | 100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375, 7975 |  |
| SWB/FB |  | 20 | 4 | 100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375 | 8000, 10000, 12000, 14000 |
|  | 21 | 3 | 100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375, 7975 | 10000, 12000, 16000 |

For each partition,corresponds to the frequency of the last band in the *i*-th partition. The indicesand  of the first and last bands in each spectral partition can be derived as a function of the core sampling rate and the FFT size:

, (1403)

, (1404)

whereis the frequency of the first band in the first spectral partition. Hence the FD-CNG generates some comfort noise above 50Hz only.

##### 5.6.3.6.1 SID parameters interpolation

The SID parameters are interpolated using linear interpolation in the log domain, as described in subclause 6.7.3.1.2. The interpolated SID parameters are noted .

##### 5.6.3.6.2 LPC estimation from the interpolated SID parameters

A set of LPC coefficients is estimated from the SID spectrum in order to update excitation and LPC related memories, as described in subclause 6.7.3.1.3. The LPC coefficients are noted .

##### 5.6.3.6.3 FD-CNG encoder comfort noise generation

A FD-CNG time-domain signal is generated similarly to the time-domain CNG signal generated at the decoder-side (see subclauses 6.7.3.3.2 and 6.7.3.3.3), except that the interpolated SID parameters  are used to generate the noise in the frequency-domain (instead of the noise levels  which are not available at the encoder side).

##### 5.6.3.6.4 FD-CNG encoder memory update

The memories update is performed using the FD-CNG time-domain signal and the LPC coefficients , similarly to the decoder memory update (see subclause 6.7.3.3.4).

Additionally, the weighted signal domain memories are updated by filtering the FD-CNG time-domain signal through a LP analysis filter with a weighted version of the LPC coefficients .