### 5.2.6 Coding of upper band for LP-based Coding Modes

#### 5.2.6.1 Bandwidth extension in time domain

The time-domain bandwidth extension (TBE) module codes the signal content beyond the range of frequencies that are coded by the low band core encoder. The inaccuracies in the representation of the spectral and the temporal information content at higher frequencies in a speech signal are masked more easily than contents at lower frequencies. Consequently, TBE manages to encode the spectral regions beyond what is coded by the core encoder in speech signals with far fewer bits than what is used by the core encoder.

The TBE algorithm is used for coding the high band of clean and noisy speech signals when the low band is coded using the ACELP core. The input time domain signal of current frame is divided into two parts: the low band signal and the super higher band (SHB) signal for SWB signal () or the higher band (HB) signal for WB signal (). While the SWB TBE encoding of upper band (6.4 to 14.4 kHz or 8 to 16 kHz) is supported at bitrates 9.6 kbps, 13.2 kbps, 16.4 kbps, 24.4 kbps, and 32 kbps; the WB TBE encoding of upper band (6 to 8 kHz) is supported at 9.6 kbps and 13.2 kbps. Specifically, at 13.2 kbps and 32 kbps the SWB TBE algorithm is employed as the bandwidth extension algorithm when the speech/music classifier determines that the current frame of signal is active speech or when GSC coding is selected to encode super wideband noisy speech. Similarly, at 9.6 kbps, 16.4 kbps and 24.4 kbps the TBE algorithm is used to perform bandwidth extension for all ACELP frames. The ACELP/MDCT core coder selection at 9.6 kbps, 16.4 kbps and 24.4 kbps is based on an open-loop decision as described in subclause 5.1.14.1.

In coding the higher frequency bands that extend beyond the narrowband frequency range, bandwidth extension techniques exploit the inherent relationship between the signal structures in these bands. Since the fine signal structure in the higher bands are closely related to that in the lower band, explicit coding of the fine structure of the high band is avoided. Instead, the fine structure is extrapolated from the low band. The high level architecture of the TBE encoder is shown in figure below. The front end processing to generate the high band target signal in figure , is replaced by a simple flip-and-decimate-by-4 operation in the WB TBE framework.



Figure 44: Time Domain Bandwidth Extension Encoder. Blocks in dashed lines are activated only at higher bit rates (24.4 kb/s or higher)

##### 5.2.6.1.1 High band target signal generation

The input speech signal, , that is sampled at either 32 kHz or 48 kHz sampling frequency, is processed through the QMF analysis filter bank. This processing step is performed as part of the common processing step as described in subclause 5.1. The output of the QMF analysis filter bank,  and  are the real and imaginary sub-band values, respectively,  is the sub-band time index with  and  is sub-band frequency band index with . The number of sub-bands, , where *Fs* is the sample rate of the input signal, . For example, for a SWB 32 kHz sampled input, *Lc* = 40, and for a FB 48 kHz sampled input, *Lc* = 60 sub-bands.

The spectral flip and down mix module extracts the high band components from the input speech signal. The high band target signal contains the high frequency components of the input signal that are to be represented by the time domain bandwidth extension encoder. The frequency range of the high band depends on the coding bandwidth of the low band ACELP core. When the low band ACELP core codes up to a maximum bandwidth of 6.4 kHz, then the high band target signal, , contains input signal components in the 6.4 -- 14.4 kHz band. When the low band ACELP core codes up to a maximum bandwidth of 8 kHz, then the high band target signal contains input signal components in the 8 -- 16 kHz band. In the high band target signal, , the input signal components are arranged in a flipped and down-mixed format such that the 0 frequency value in the SWB high band target signal corresponds to the maximum frequency in the above mentioned bandwidth. The process of deriving the SWB high band target signal is shown pictorially in figure .



Figure 45: Generation of high band target signal for SWB TBE

Similarly, the WB high band target signal containing signal components from 6 to 8 kHz is generated as shown in figure . In particular, a simple flip-and-decimate-by-4 operation is performed to extract the 6 to 8 kHz high band from the 16 kHz sampled WB input signal. In case of fixed point operation, the input signal is scaled dynamically based on the spectral tilt of the input signal prior to performing the flip-and-decimate-by-4 operation. This adjustment of the scaling is done such that for signals with low energy in the high band (indicated by having a spectral tilt >= 0.95), no headroom is provided prior to decimation to avoid any loss of precision after decimation, and for signals with relatively higher energy in the high band (indicated by having a spectral tilt < 0.95), a headroom of 3 bits is provided prior to decimation to avoid any saturation in the decimation operation.



Figure 46: Generation of high band target signal for WB TBE

The process of deriving the SWB high band target signal using the complex low-delay filter bank (CLDFB) analysis is described below. The real, , and imaginary, , CLDFB coefficients are first flipped as follows:

 ()

The values *m* and *M* are dependent on the maximum bandwidth coded by the ACELP core as shown in table . The flipped coefficients,  and are used to generate the high band target signal using the CLDFB synthesis as described in subclause 6.9.3.

Table 52: Maximum and minimum frequencies for spectral flipping and downmix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ACELP core | | m | M | High band target signal band |
| Low band core sample rate | | Low band coded bandwidth |  |
| @12.8 kHz core | 6.4 kHz | 14 | 34 | 6.4-14.4 kHz |
| @ 16 kHz core | 8 kHz | 19 | 39 | 8-16 kHz |
| @12.8 kHz core | 6.4 kHz | - | - | 6-8 kHz (WB) |

##### 5.2.6.1.2 TBE LP analysis

TBE linear prediction analysis is performed on a 33.75ms high band signal that includes the current 20ms SHB target frame with 5ms of past samples and 8.75ms of look ahead samples. The SWB high band target signal  is generated using the CLDFB synthesis filter bank described in subclause 6.9.3. Twenty milliseconds of the high band target signal is used to populate the shb\_old\_speech buffer from sample 220 to sample 539 as shown in figure . Both the WB TBE and SWB TBE LP analysis follow similar steps as described below except that the buffer lengths in SWB TBE and WB TBE are different reflecting the effective upper band coding bandwidth. Certain steps that are relevant at low bit rates and only for WB TBE and not for SWB TBE LP analysis are specified where applicable.



Figure 47: SWB Highband target signal buffers

The first 220 samples in the shb\_old\_speech buffer are filled from the memory shb\_old\_speech\_mem. The shb\_old\_speech\_mem is updated as

 ()

For linear prediction analysis, a 540 sample LPC analysis frame is derived from the shb\_old\_speech buffer. A single analysis window , is used to calculate 10 auto correlation coefficients. The window function is as shown in figure



Figure 48: SWB TBE analysis window

The autocorrelation coefficients  is calculated according to

 ()

where  is obtained by multiplying by .

A bandwidth expansion is applied to the autocorrelation coefficients by multiplying the coefficients by the expansion function:

 ()

The bandwidth expanded autocorrelation coefficients are used to obtain LP filter coefficients,  by solving the following set of equations using the Levinson-Durbin algorithm.

 ()

It should be noted that 

##### 5.2.6.1.3 Quantization of linear prediction parameters

First a spectral bandwidth expansion operation is performed on the LPC coefficients by multiplying  with bandwidth expansion weights 

 ()

The  coefficients are given in table :

Table 53: LPC bandwidth expansion coefficients

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k | | | | | | | | | | |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | |
| 0.975 | 0.950625 | 0.926859375 | 0.903687891 | 0.881095693 | 0.859068301 | 0.837591593f | 0.816651804 | 0.796235509 | 0.776329621 | |

The bandwidth expansion of LP coefficients is performed in WB TBE LP analysis and low bitrate SWB TBE analysis for bit rates below 13.2 kbps. The bandwidth expanded LP coefficients  are converted to line spectral frequencies 

The LSP weights  are calculated from the line spectral frequencies 

 ()

And based on the spacing between adjacent LSF parameters,  and 

 ()

 ()

where  and , otherwise.

5.2.6.1.3.1 LSF quantization

WB TBE LSF quantization uses a single stage vector quantizer that utilizes 2 bits and 8 bits for LSF quantization, respectively, at 9.6 kbps and at 13.2 kbps. For low bit rate SWB LSF quantization at 9.6 kbps and primary frame encoding in channel aware mode at 13.2 kbps (see subclause 5.8.3.1), a simple 8-bit 10-dimensional single stage vector quantizer is used. In SWB TBE encoding, for bit rates at and above 13.2 kbps, the 10 dimensional LSF vectoris encoded in a SQ and extrapolation procedure. The low-frequency half (first five LSF coefficents ) of the LSF vector are scalar quantized with the following number of bits {4, 4, 3, 3, 3}. That is the first two coefficents are quantized with four bits each, with the CB from table , while the remaining three coefficients are quantized with three bits each with reconstruction values in table .

Table 54: Four bit CB for LSF quantization

|  |  |
| --- | --- |
| Index | Value |
| 0000 | 0.01798018 |
| 0001 | 0.02359377 |
| 0010 | 0.02790103 |
| 0011 | 0.03181538 |
| 0100 | 0.03579450 |
| 0101 | 0.03974377 |
| 0110 | 0.04364637 |
| 0111 | 0.04754591 |
| 1000 | 0.05181858 |
| 1001 | 0.05624165 |
| 1010 | 0.06022101 |
| 1011 | 0.06419064 |
| 1100 | 0.06889389 |
| 1101 | 0.07539274 |
| 1110 | 0.08504436 |
| 1111 | 0.10014875 |

Table 55: Three bit CB for LSF quantization

|  |  |
| --- | --- |
| Index | Value |
| 000 | 0.02070812 |
| 001 | 0.02978384 |
| 010 | 0.03800822 |
| 011 | 0.04548685 |
| 100 | 0.05307309 |
| 101 | 0.06137543 |
| 110 | 0.07216742 |
| 111 | 0.09013262 |

The output of the SQ is five quantized coefficients , which are also used as an input to extrapolation of the remaining five coefficients.

The high-frequency half (last five LSF coefficents ) are calculated as weighted averaging of the quantized low-frequency part of the LSF vector and selected optimal grid points. First the already quantized coefficents are flipped arround a quantized mirroring frequency, which separates the low-frequency part from the high-frequency part of the LSF vector. Then the the fipped coefficients are adjusted by an optimal frequency grid codebook, which is obtained in a closed-loop search procedure.

The transmitted mirroring frequency  is obtained as quantized difference between the first extrapolated and last quantized coefficient, added to the last coded position

 ()

The quantization is performed with two bit SQ with reconstruction points {0.01436178, 0.02111641, 0.02735687, 0.03712105}.

The quanized five low-frequency LSF coefficents  are flipped arround the mirroring frequency , to form a set of flipped coefficients 

 ()

Then the flipped coefficients are further scaled according to

 ()

The flip and scaled coefficients are adjusted with one of four grid vectors. The pre-stored set of four grid vectors from table , is first re-scaled to fit into the interval between the last quantized LSF  and maximum grid point value 0.5

 ()

Table 56: CB with LSF grid points

| grid 1 | grid 2 | grid 3 | grid 4 |
| --- | --- | --- | --- |
| 0.15998503 | 0.15614473 | 0.14185823 | 0.15416561 |
| 0.31215086 | 0.30697672 | 0.26648724 | 0.27238427 |
| 0.47349756 | 0.45619822 | 0.39740108 | 0.39376780 |
| 0.66540429 | 0.62493785 | 0.55685745 | 0.59287916 |
| 0.84043882 | 0.77798001 | 0.74688616 | 0.86613986 |

Then smoothed coefficients  are obtained by weighed averaging of the re-scaled grid sets  and re-scaled flipped coefficients 

 ()

where  are pre-defined weights 

Different grid vectors  form different sets of smoothed coefficients . The optimal grid is selected as the one that produces smoothed coefficients  closest, in the mean-squared-error sense, to the actual high-frequency coefficients 

 ()

The CB indices for the five low-frequency coefficients, mirroring frequency index, and the index of the optimal grid are transmitted to the decoder.

The quantized LSF parameters  are checked for inter-LSF spacing. First, the spacing between adjacent LSFs is determined

 ()

If , then the quantized LSFs are adjusted as follows

If , then

.

If , then

.

Otherwise,



##### 5.2.6.1.4 Interpolation of LSF coefficients

The quantized LSF parameters  are converted into LSPs 

 ()

The interpolation between the LSPs of the current frame, , and that of the previous frame  is performed to a set of 4 interpolated LSPs,  as follows:

 ()

The interpolated LSPs are then converted back to LSFs. This process gives 4 sets of interpolated LSFs. Further the LSFs are converted into LP coefficients . Note  for all *j*. The conversion from the interpolated LSFs to LP coefficients is done similar to the process employed in the core coding modules.

5.2.6.1.4.1 LSP Interpolation at 13.2 kbps and 16.4 kbps

LSP interpolation is employed to smooth the SWB TBE LSP parameters at 13.2 kbps and 16.4 kbps. When the bit rate of operation is 13.2 kbps or 16.4 kbps, the similarity of signal characteristics of the current frame and the previous frame are analysed to determine whether they meet the Preset Modification Conditions (PMCs) or not. The PMCs denote if the PMCs are met, then a first set of correction weights are determined from the Linear Spectral Frequency (LSF) differences of the current frame and the LSF differences of the previous frame. Otherwise a second set of correction weights are set with constant values that lie in the range 0.0 to 1.0. The LSPs are then interpolated using the appropriate set of correction weights. The PMCs are used to determine whether the signal characteristic of the current frame and the previous frame is close or not.

The PMCs are determined by first converting the linear prediction coefficients (LPCs) to reflection coefficients (RCs) as follows using a backwards Levinson Durbin recursion. For a given N-th order LPC vector the Nth reflection coefficient value is derived using the equality, it is then possible to calculate the lower order LPC vectors using the following recursion.

 ()

which yields the reflection coefficient vector .

A spectral tilt parameter  is then calculated from the first reflection coefficient  as follows:

 ()

The PMCs are then determined from the spectral tilts of the current frame and the previous frame along with the coding types of the current and the previous frames. In the case that the current frame is a transition frame, the PMCs are by default not met.

In order to determine that the current frame is not a transition frame requires the following steps:

1. Determining whether there is a change from a fricative frame to a non-fricative frame. Specifically, if the tilt of the previous frame is greater than a tilt threshold value of 5.0 and the coder type of the frame is a transient, or the tilt of the previous frame is greater than a tilt threshold value of 5.0 and the tilt of the current frame is smaller than a tilt threshold value of 1.0 then the frame is a transition frame.
2. Determining whether there is a change from a non-fricative frame to a fricative frame. Specifically, if the tilt of the previous frame is smaller than a tilt threshold value of 3.0 and the coder type of the previous frame is equal to one of the four types of VOICED, GENERIC, TRANSITION or AUDIO, and the tilt of the current frame is greater than a tilt threshold value of 5.0 then again the frame is a transition frame
3. The current frame is not a transition frame if both 1) and 2) are not met, and then the PMCs are met.

When the PMCs are met, the first set of correction weights are defined as follows

 ()

where the  and  are defined as

 ()

 ()

where the are the LSFs of the current frame and the are the LSFs of the previous frame.

When the PMCs are not met, the second set of correction weights is used and these are set to 0.5 as follows:

 ()

 ()

The correction weights are finally used in the interpolation between the LSPs of the current frame and the LSPs of the previous frame, it is described as follows:

 ()

where  are the LSPs of the previous frame, are the LSPs of the current frame, are the interpolated LSPs. The interpolated LSPs are used to encode the current frame.

The interpolated LSPs are then converted back to LSFs，further the LSFs are converted into LP coefficients.

5.2.6.1.4.2 LSP Interpolation at 24.4 kbps and 32 kbps

The interpolation between the LSPs of the current frame, , and that of the previous frame  is performed to yield a set of 4 interpolated LSPs,  as follows:

 ()

Table 57: LSP interpolation factors IC1 and IC2 over four sets

| LSP set, j | IC1 | IC2 |
| --- | --- | --- |
| Set 1 | 0.7 | 0.3 |
| Set 2 | 0.4 | 0.6 |
| Set 3 | 0.1 | 0.9 |
| Set 4 | 0 | 1.0 |

The interpolated LSPs are then converted back to LSFs. This process gives 4 sets of interpolated LSFs. Further the LSFs are converted into LP coefficients. Note  for all *j*.

##### 5.2.6.1.5 Target and residual energy calculation and quantization

At 24.4 kb/s and 32 kb/s, energy parameters from the high band target frame and an LPC residual calculated from the target signal and the interpolated LP coefficients, , are calculated and transmitted to the decoder.

For calculation of the energy parameters, the signal in the high band target frame (see figure ), denoted as  is used. The 4 energy parameters are calculated according to:

 ()

The sum of energy parameters, , is quantized using 6 bits in SWB TBE at 24.4 kbps and 32 kbps. A residual signal  is calculated from  using the interpolated LP coefficients. For j=1, …4 and n=0,…79,

 ()

The residual signal is then used to calculate the residual energy parameters,

 ()

The residual energy parameters are then normalized. First the maximum residual energy parameter is calculated  . Then  is normalized to get  as per

 ()

Each is then scalar quantized using a 3 bit uniform quantizer with the lowest quantization point as 0.125 and uniform quantization steps of 0.125.

##### 5.2.6.1.6 Generation of the upsampled version of the lowband excitation

An upsampled version of the low band excitation signal is derived from the ACELP core as show in figure below.



Figure 49: Generating the upsampled version of the lowband excitation

For each ACELP core coding subframe, *i*, a random noise scaled by a factor voice factor,  is first added to the fixed codebook excitation that is generated by the ACELP core encoder. The voice factor is determined using the subframe maximum normalized correlation parameter,  that is derived during the ACELP encoding. First the  factors are combined to generate .

 ()

 calculated above is limited to a maximum of 1 and a minimum of 0.

 if the ACELP core encodes a maximum of 6.4 KHz or  if the ACELP core encodes a maximum bandwidth of 8 KHz.

The resampled output is scaled by the ACELP fixed codebook gain and added to a delayed version of itself.

 (696)

where gc is the subframe ACELP fixed codebook gain, gp is the subframe ACELP adaptive codebook gain and P is the open loop pitch lag.

##### 5.2.6.1.7 Non-Linear Excitation Generation

The excitation signal  is processed through a non-linear function in order to extend the pitch harmonics in the low band signal into the high band. The non-linear processing is applied to a frame of  in two stages; the first stage works on the first half subframe (160 samples) of  and the second stage works on the second half subframe. The non-linear processing steps for the two stages are described below. In the first stage , and in the second stage, .

First, the maximum amplitude sample  and its location relative to the first sample in the stage  are determined.

 ()

Based on the value of , the scale factor  is determined.

 ()

The scale factor  and the previous scale factor parameter from the memory  are then used to determine the parameter scale step.

 ()

If , then 

The output of the non-linear processing  is derived as per

 ()

If the current frame is VOICED frame and sum of voice factors over the subframes is less than a threshold, , then the sign reversal when  is not performed. The threshold, = 0.70 when there are five subframes per frame and = 0.78 when there are four subframes per frame.

The previous scale factor parameter is updated recursively for all  according to

for(j=n1; j< n2; j++)

if(j<imax)

=

end

end

##### 5.2.6.1.8 Spectral flip of non-linear excitation in time domain

The non-linear excitation  is spectrally flipped so that the high band portion of the excitation is modulated down to the low frequency region. This spectral flip is accomplished in time domain

 ()

##### 5.2.6.1.9 Down-sample using all-pass filters

 is then decimated using a pair of all pass filters to obtain an 8 kHz bandwidth (16 kHz sampled) excitation signal . This is done by filtering the even samples of  by an all pass filter whose transfer function is given by

 ()

And the odd samples of  by an all pass filter whose transfer function is given by

 ()

The 16 kHz sampled excitation signal  are obtained by averaging the outputs of the above filter.

These filter coefficients are specified in below.

Table 58: All-pass filter coefficients for decimation by a factor of 2

|  | All pass coefficients |
| --- | --- |
| a0,1 | 0.06056541924291 |
| a1,1 | 0.42943401549235 |
| a2,1 | 0.80873048306552 |
| a0,2 | 0.22063024829630 |
| a1,2 | 0.63593943961708 |
| a2,2 | 0.94151583095682 |

##### 5.2.6.1.10 Adaptive spectral whitening

Due to the nonlinear processing applied to obtain the excitation signal , the spectrum of this excitation is no longer flat. In order to flatten the spectrum of the excitation signal , 4th order linear prediction coefficients are estimated from The spectrum of  is then flattened by inverse filtering  using the linear prediction filter.

The first step in the adaptive whitening process is to estimate the autocorrelation of the excitation signal

 ()

A bandwidth expansion is applied to the autocorrelation coefficients by multiplying the coefficients by the expansion function:

 ()

The bandwidth expanded autocorrelation coefficients are used to obtain LP filter coefficients,  by solving the following set of equations using the Levinson-Durbin algorithm as described in section.

 ()

It must be noted that .

The whitened excitation signal  is obtained from  by inverse filtering

 ()

4 samples of  from the previous frame are used as memory for the above filtering operation.

For bit rates 24.4 kb/s and 32 kb/s, the whitened excitation is further modulated (in a two-stage gain shape modulation) by the normalized residual energy parameter. In other words, for bitrates 24.4 kb/s and 32 kb/s,

 ()

##### 5.2.6.1.11 Envelope modulated noise mixing

To the whitened excitation, a random noise vector whose amplitude has been modulated by the envelope of the whitened excitation is mixed using a mixing ratio that is dependent on the extent of voicing in the low band.

First,  is calculated and then the envelope of the envelope of the whitened excitation signal is calculated by smoothing 

 ()

In SWB mode, the factors  and  are calculated using the voicing factors,  for subframes , which calculated by parameters which determined from the low band ACELP encoder. The voicing factors denote voicing extend of the high band signal since the fine signal structure in the higher bands are closely related to that in the lower band. The average of the 4 voicing factors, , is calculated and modified as . This is then confined to values between 0.6 and 0.999. Then  and  are estimated as

 ()

 ()

However, for bit rates 16.4 kb/s and 24.4 kb/s and if TBE was not used in the previous frame,  and  are set to

 ()

 ()

and for , is substituted by an approximated value  as

 ()

In WB mode, the factors  and are initialized to and. However, if the bitrate is 9.6kb/s, they are reset to andif the low band coder type is voiced or, or to and if the low band coder type is unvoiced or. In SWB mode, for bit rates 24.4 kbps and 32 kbps, the mix factors are estimated based on both the low band voice factor, , and the high band closed-loop estimation, where



The  mix factor is estimated based on the high band residual signal, , transformed low band excitation, , and the modulated white noise excitation .

A vector of random numbers,  of length 160 is then modulated by  to generate  as

 ()

The whitened excitation is then de-emphasized with  which is the pre-emphasised effect since the used spectrum is flipped.

 ()

If the lowband coder type is unvoiced, the excitation  is first rescaled to match the energy level of the whitened excitation 

 ()

where

 ()

And then pre-emphasised with =0.68 to generate the final excitation which is the de-emphasised effect since the used spectrum is flipped.

 ()

If the lowband coder type was not un-voiced, the final excitation is calculated as

 (720)

for each sample index  within subframe .

For bit rates less than 24.4 kb/s, the mixing parameters  and  are estimated as,

 ()

 ()

For bit rates 24.4 kb/s and 32 kb/s, the mixing parameters , are estimated for as follows:

 (723)

 (723a)

where the parameter  is defined in equation ().

 is then de-emphasised to generate the final excitation.

##### 5.2.6.1.12 Spectral shaping of the noise added excitation

The excitation signal  is then put through the high band LPC synthesis filter that is derived from the quantized LPC coefficients (see subclause 5.2.4.1.3).

For bitrates below 24.4 kb/s, a single LPC synthesis filter  is used and the shaped excitation signal  is generated as

 ()

For bitrates at and above 24.4 kb/s the LPC synthesis filter is applied to the excitation signal  in four subframes based on

 ()

In particular, for bit rates at and above 24.4 kbps, first a memory-less synthesis is performed (with past LP filter memories set to zero) and the energy of the synthesized high band is matched to that of the original signal. In the subsequent step, the scaled or energy compensated excitation signal as shown below, , is used to perform synthsesis in the second step.



##### 5.2.6.1.13 Post processing of the shaped excitation

The shaped excitation is the synthesized high band signal which is generated by passing the excitation signal  through the LPC synthesis filter. The excitation signal  is determined by the low band model parameters and the coefficients of the LPC synthesis filter are determined by the high band model parameters. A short-term post-filter is applied to the synthesized high band signal to obtain a short-term post-filtered signal. Comparing with the shape of the spectral envelope of the synthesized high frequency band signal, the shape of the spectral envelope of the short-term post-filtered signal is closer to the shape of the spectral envelope of the high-frequency band signal. The short-term post-filter includes a pole-zero filter, the coefficients of the pole-zero filter are set by the set of high band model parameters. It is described as follows:

 ()

 ()

 is derived from the quantized LPC coefficients (see subclause 5.2.6.1.3) with the factors  and controlling the degree of the short-term post‑filtering.  are the quantized LPC coefficients. The factors and are calculated according to:

 ()

where  is parameter that jointly controls envelope shape and excitation noisiness. It is based on the spectral tilt, determined by the LPC coefficient:

 ()

where  is the past value of  .

 ()

 (731)

The gain term  is calculated from the truncated impulse response  of the filter and is given by:

 ()

The shaped excitation  is divided into four subframes, and each subframe is filtered through , and then filtered with the synthesis filter  to produce .

After filtering the synthesized high band signal using the pole-zero filter, filter then compensates for the tilt in the pole-zero filter  and is given by:

 ()

 ()

where is set to default constant value or adaptively calculated according to the high band coding parameters and the synthesized high band signal. The calculation process is as follows:  is a tilt factor, with  being the first reflection coefficient calculated from by:

 ()

A gain term  is applied to compensate for the decreasing effect of in. It has been shown that the product filter has a gain close to unity. is set according to the sign of. If  is positive,, otherwise,.

Then is passed through the tilt compensation filter  resulting in the post‑filtered speech signal 

Adaptive Gain Control (AGC) is applied to compensate for any gain difference between the synthesized speech signal and the post‑filtered signal. The gain scaling factor  for each subframe is calculated by:

 ()

and the post processed shaped excitation is given by:

 ()

where  is updated in sample‑by‑sample basis and given by:

 ()

and where  is an AGC factor with value of 0.85.

In order to smooth the evolution of the post-processed spectrally shaped highband excitation signal across frame boundaries, the look-ahead and the overlap samples are scaled based on the ratio of the current frame’s energy in the overlap region and the previous frame’s energy in the overlap region. The scale factor computation is performed as shown in equation (1579) in subclause 6.1.5.1.12.

The tenth-order LPC synthesis performed as described according to subclause 5.2.6.1.12 uses a memory of ten samples, thus there is at least an energy propagation over ten samples from the previous frame into the current frame. When calculating the energy scaling to be applied to the current frame, the first 10 samples of the current frame are considered as a part of previous frame energy. If the voicing factor is greater than 0.75, the numerator in equation (1579) is attenuated by 0.25. The spectrally shaped high band signal is then modified by the scale factor as shown in equation (1580) in Clause 6.1.5.1.12.

##### 5.2.6.1.14 Estimation of temporal gain shape parameters

There are different initialization estimation of temporal gain shape for the WB mode and the SWB mode.

5.2.6.1.14.1 Initialization estimation of temporal gain shape for WB mode

The frame is divided into eight segments, and the energy envelope of each segment is calculated, then the 4 gain shapes are calculated from the calculated 8 energy envelopes.

The energy envelopes of the eight segments of the target signal and the shaped excitation signal are calculated as follows:

Asymmetric windows are applied to the first and the eighth segments,

 ()

 ()

And symmetric windows are applied to the segments from the second to the seventh.

 ()

 ()

Four high band gain shapes are then calculated by combining pairs of energy envelopes as follows

 ()

The asymmetric windows are determined by the number of look head samples and the number of segments. The asymmetric window and the symmetric windows are described as follows, the number of look-ahead samples is 5.



Figure 50: Asymmetric window and symmetric window

The variable  for is tabulated in table below:

Table 59: Window for highband gain shape calculation

| Index | Value |
| --- | --- |
| 0 | 0.0 |
| 1 | 0.15643448 |
| 2 | 0.30901700 |
| 3 | 0.45399052 |
| 4 | 0.58778524 |
| 5 | 0.70710677 |
| 6 | 0.80901700 |
| 7 | 0.89100653 |
| 8 | 0.95105654 |
| 9 | 0.98768836 |
| 10 | 1.0 |

5.2.6.1.14.2 Initialization estimation of temporal gain shape for SWB mode

The high band target frame (see subclause 5.2.6.1.1),  and the post-processed shaped excitation  are used to calculate 4 temporal gain shape parameters,  for. The gain shape parameters are calculated using an overlap of 20 samples from the previous frame to avoid transition artifacts during the reconstruction at the decoder.

 (744)

where the subframe energies in the target high band signal and the shaped excitation signal are calculated as

 (745)

and

. (746)

The window function  is a window signal is given by

 (747)

The variable  for  is tabulated in table below.

Table 60: Window for highband gain shape calculation

| Index | Value |
| --- | --- |
| 0 | 0.0 |
| 1 | 0.006156 |
| 2 | 0.024472 |
| 3 | 0.054497 |
| 4 | 0.095492 |
| 5 | 0.146447 |
| 6 | 0.206107 |
| 7 | 0.273005 |
| 8 | 0.345492 |
| 9 | 0.421783 |
| 10 | 0.5 |
| 11 | 0.578217 |
| 12 | 0.654508 |
| 13 | 0.726995 |
| 14 | 0.793893 |
| 15 | 0.853553 |
| 16 | 0.904508 |
| 17 | 0.945503 |
| 18 | 0.975528 |
| 19 | 0.993844 |
| 20 | 1.0 |

5.2.6.1.14.3 Additional processing of temporal gain shape

Additionally, the gain shape values are normalized such that

 ()

The 4th subframe gain shape value from the last subframe  is used to smooth the Gain Shape parameter evolution. A variable  is calculated as

 ()

If the sum of the voicing factors , then the gain shape parameters are smoothed as follows:

 ()

 ()

The gain shape parameters vector corresponding to a given frame are transformed into the log domain and vector quantized. The quantized gain shape parameters are denoted 

##### 5.2.6.1.15 Estimation of frame gain parameters

In addition to the gain shape parameter, an overall frame gain parameter  is calculated. First the spectrally shaped excitation is scaled by the gain shape parameters.

 ()

Samples with negative indices are obtained from the previous frame and

 ()

The energies of the  and  for the entire duration of the frame and the overlaps is calculated using a window:

 ()

 ()

where the negative samples are obtained from previous frames, and the window function  is given by

 ()

where  is defined in table .

If the high band target energy as calculated in Equation (754) is saturated, then the target signal, , is scaled based on the number of subframes that are saturated (from clause 5.2.6.1.14.2). Then the high band target energy is recalculated using the scaled target signal using Equation (754). Subsequently, the gain frame GF in Equation (757) is compensated for the scaling performed on the target signal.

The overall frame gain parameter  is calculated as

 ()

The parameter  is attenuated if the quantization of the gain shape parameter is poor. First the power-off-the-peak value for unquantized and quantized gain shape parameters is calculated:

 ()

 ()

If , then the gain shape parameter quantization is deemed poorly quantized and the gain frame parameter is attenuated as follows in order to avoid perceptually annoying artifacts in the reconstructed speech signal at the decoder.

 ()

Further, the parameter is smoothed if the current frame is determined to be similar in spectral characteristics to the previous frames. To determine similarity in spectral characteristics, the fast and slow evolution rates,  and  , and the minimum LSP spacing  are determined as follows:

 ()

where  and  andwhen k=1 and otherwise.

A smoothed version of  using thevalues from the previous frames is also determined:

 ()

If the minimum LSP spacing and smoothed LSP spacing satisy a spacing critierion (e.g., *δmin* < 0.008, *δsmoothed* < 0.005) indicating an artifact generating condition, the high band target is filtered using the high band LP to attenuate the coding artifacts before estimating the gain shape and gain frame parameters.

If the  and the values are smaller than 0.001, then the gain frame parameter  is smoothed between the previous and the current frame.

 ()

Also, if, then an additional attenuation of the  is performed: . Also, if the current frame’s , then  is modified as per

 ()

where  is calculated as described in subclause 5.2.6.1.15.

For the WB mode and if the bitrate is 9.6 kb/s, the gain frame parameter is further adapted, based on the average of the 4 voicing factors, , as , if the low band coder type is voiced, and as , if the low band coder type is not voiced, but .

The gain frame parameter  is scalar quantized in the log domain using 5 bits. The lowest quantization point is chosen to be -1.0 and the quantization steps are set to be 0.15.

##### 5.2.6.1.16 Estimation of TEC/TFA envelope parameters

The Temporal Envelope Coding (TEC) and the Temporal Flatness Adjuster (TFA) shape the temporal envelope of the high frequency band signal generated by the TBE. They are enabled as the post processing of the TBE at 16.4 and 24.4 kbps with the output sampling frequencies higher than 8000 Hz.

The TEC is for getting better shapes of onsets at the decoder by using the temporal envelope of the low frequency band and the transmitted information on the temporal envelope of the high frequency band. The information indicates the two different shapes, one is “steep onset” and the other is “gentle onset”. The TEC has two modes for accommodating those shapes of the onsets, “no smoothing mode” is for steep onsets and “smoothing mode” is for gentle onsets. Thus, the transmitted information represents the mode of TEC to be used at the decoder as well as the shape of the onset. At the decoder, the information is used to calculate the temporal envelope of the high frequency band.

The TFA flattens the temporal envelope of the high frequency band according to the transmitted information on the flatness of the temporal envelope of the input signal.

5.2.6.1.16.1 Estimation of TEC parameters

The TEC parameter  to be transmitted to the decoder is estimated as follows, which is the information on the temporal envelope of the high frequency band and indicates the activation and the used mode of the TEC at the decoder.

The temporal envelope of the low frequency band of the input signal is defined as the mean of the temporal envelopes of three sub-bands in the low frequency band (sub-low-frequency-bands) in the CLDFB domain. The temporal envelope of the *m*-th sub-low-frequency-band is calculated by

 ()

where  is the energy in the CLDFB domain described in subclause 5.1.2.2 and,  and  are the lower and upper limits of the CLDFB sub-band of the *m*-th sub-low-frequency-band. And then, the temporal envelope of the low frequency band  is calculated by

 ()

The temporal envelope of the high frequency band of the input signal is calculated in the CLDFB domain.

 ()

where  is the energy in the CLDFB domain described in subclause 5.1.2.2 and,  and  are the lower and upper limits of the CLDFB sub-band in the frequency range of the TBE.

The shape of the onset is detected by these temporal envelopes of the high and low frequency bands. A steep onset is detected and the no smoothing mode is selected when . And, a gentle onset is detected and the smoothing mode is selected when .

 ()

where  and  are the variances of and respectively.

When , the maximum values of the temporal envelopes of the high and low frequency bands and their positions are detected:

 ()

 ()

And then, the local minimum and maximum values of the temporal envelope of the high frequency band at the position *i*, are detected and the difference between the local maximum and minimum values is calculated:

 ()

 ()

The maximum of the difference and its position are detected:

 ()

Using the parameters above, is set as

 ()

When , the smoothed temporal envelope of the low frequency band is calculated

 ()

where

 ()

And then, the correlation coefficient between  and  is calculated. Further, the ratio between the variances of  and is calculated:

 ()

Using these parameters, is set as

 ()

5.2.6.1.16.2 Estimation of TFA parameters

The parameter  to be transmitted to the decoder is estimated as follows, which represents the flatness of the temporal envelope of the high frequency band as well as the activation of the TFA at the decoder.

The flatness measure of the temporal envelope of the signal in the high band target frame  is calculated as:

 ()

where and where .

In addition to the flatness measure, the mean of the pitch lags and the mean of the open-loop pitch gains of the half frames are calculated. Finally, depending to the last core mode, the parameter is estimated by the parameters above:

in the case that the last core mode is not TCX20,

 ()

in the case that the last core mode is TCX20,

 ()

5.2.6.1.16.3 Set the transmitted parameter for TEC and TFA

The TEC and TFA parameters are jointly coded by 2 bits and then transmitted to the decoder together with the other TBE information:

 ()

##### 5.2.6.1.17 Estimation of full-band frame energy parameters

The full-band TBE algorithm is used for coding the full band. Since the fine signal structure in the full bands are closely related to that in the high band and low band, the full band of the signal are coded by using only 4 bits together with the information from the low band and the high band. A synthesized full band signal is obtained by coding the high band and predicting spectrum from the high band, the synthesized full band signal is then de-emphasized with which is determined by the characteristic factors derived from coding the low band signal. A band passed full band signal is obtained by band-pass filtering the input signal. The energy ratio is calculated by comparing the energy calculated from the de-emphasized synthesized full band signal with the energy calculated from the band passed full band signal. Finally the parameters including the characteristic factors, high band coding information and the energy ratio are transmitted to the decoder.

The synthesized full band signal is obtained as follow: A vector of random numbers of length 320 passes through the LPC synthesis filter, the spectrum of  is used as the predicted spectrum from the high band and the coefficients of the LPC synthesis filter are derived from the quantized LPC coefficients (see subclause 5.2.6.1.3).

 ()

Then the synthesized full band signal is de-emphasized as follows:

The spectrum of  are moving corrected described as follows:

 ()

The variables are the parameters of spectrum moving correction tabulated in Table below:

Table 61: Parameters of spectrum moving correction

| Index | Value | Index | Value |
| --- | --- | --- | --- |
| 0 | 9.536743164062500e-007 | 16 | -9.536743164062500e-007 |
| 1 | 9.353497034680913e-007 | 17 | -9.353497034680913e-007 |
| 2 | 8.810801546133007e-007 | 18 | -8.810801546133007e-007 |
| 3 | 7.929511980364623e-007 | 19 | -7.929511411930434e-007 |
| 4 | 7.929511980364623e-007 | 20 | -6.743495077898842e-007 |
| 5 | 5.298330165715015e-007 | 21 | -5.298329597280826e-007 |
| 6 | 3.649553264040151e-007 | 22 | -3.649552411388868e-007 |
| 7 | 1.860525884467279e-007 | 23 | -1.860525173924543e-007 |
| 8 | 0.000000000000000e+000 | 24 | 0.000000000000000e+000 |
| 9 | -1.860526737118562e-007 | 25 | 1.860527589769845e-007 |
| 10 | -3.649554116691434e-007 | 26 | 3.649554969342717e-007 |
| 11 | -5.298331302583392e-007 | 27 | 5.298331871017581e-007 |
| 12 | -6.743496214767220e-007 | 28 | 6.743496783201408e-007 |
| 13 | -7.929512548798812e-007 | 29 | 7.929513117233000e-007 |
| 14 | -8.810802114567196e-007 | 32 | 8.810802683001384e-007 |
| 15 | -9.353497603115102e-007 | 31 | 9.353497603115102e-007 |

And then flip the spectrum of  to get 

 ()

The signalis then de-emphasized with described as follows:

 ()

The de-emphasized factor  is determined by the characteristic factors of the signal such as “voicing factors”, “spectral tilt”, “short-term average energy”, “short-term average zero crossing rate”. The calculation of de-emphasized factor  using the voicing factors is described as follows:

 ()

 ()

where is the voicing factor of the ith subframe.

The signalis modulated by the gain shape values (see subclause 5.2.6.1.14) and the overall frame gain parameter (see subclause 5.2.6.1.15).

 ()

The energy of  is described as follows:

 ()

The original input signal  pass through the band-pass filter to get the band passed full band signal, and then calculate the energy  of from 16 kHz to 20 kHz. The energy ratio is calculated as follows:

 ()

 ()

 ()

The energy ratio is then transmitted to the decoder using 4 bits.

#### 5.2.6.2 Multi-mode FD Bandwidth Extension Coding

The input signal of current frame is divided into two parts: the low band signal and the super higher band (SHB) signal for SWB signal or the higher band (HB) signal for WB signal. The low band signal of the input signal is coded by LP coding modes, and the SHB or HB signal of the input signal is coded by the multi-mode FD bandwidth extension (BWE) algorithm. A classification decision process of the SHB or HB signal of input signal is first performed. Then, the multi-mode BWE algorithm for LP-based coding modes uses a combination of adaptive spectral envelope and time envelope coding for super wideband extension, and spectral envelope coding for wideband extension, according to the result of the classification decision process. The coded bitstream of low band signal, the adaptive coded bitstream of SHB or HB signal as well as the result of the classification decision process of current input signal are output. Table describes the multi-mode FD BWE at the different bitrates of operation. Theoretically, the delay of the synthesized output signal can be determined adaptively in the range of  according to the delay of the core coding algorithm and the delay of the multi-mode bandwidth extension algorithm. To achieve lowest delay for bandwidth extension coding, the super higher band (SHB) signal is delayed by . Here  is larger than . Then, the achieved lowest delay of the multi-mode bandwidth extension is  in encoder side.

Table 62: Multi-mode FD BWE at different bitrates

|  |  |  |
| --- | --- | --- |
| Bitrate [kbps] | Bandwidth | Multi-mode FD BWE |
| 7.2, 8 | WB | Blind, HARMONIC/NORMAL |
| 13.2 | WB | Guided, HARMONIC/NORMAL |
| SWB | Guided, TRANSIENT/HARMONIC/NORMAL/NOISE |
| 32 | SWB/FB | Guided, TRANSIENT/HARMONIC/NORMAL/NOISE |

##### 5.2.6.2.1 SWB/FB Multi-mode FD Bandwidth Extension

For frames declared as TRANSIENT (TS) frames or as non-TRANSIENT, a bit budget of 31 bits is allocated to the SWB Multi-mode FD Bandwidth Extension. If the super higher band (SHB) signal of the input in the previous frame or in the current frame is detected as TRANSIENT, then the current frame is also classified as TRANSIENT. Non-TRANSIENT frames can be further classified as HARMONIC (HM), NORMAL (NM) or NOISE (NS) depending upon the frequency fluctuation that is detected. Two bits of the bit budget are allocated to the signal class. In case of a TRANSIENT frame, the remaining 29 bits are allocated to encode four spectral envelopes and four time envelopes. For other cases, i.e. non-TRANSIENT frames, the remaining 29 bits are allocated to encode fourteen spectral envelopes, and no time envelope is encoded. For FD BWE encoding, the 320 MDCT coefficients of the SHB signal,  are coded. In the case of the FB mode, the encoding algorithm from 8kHz to 15.5kHz is the same as the SWB mode. To encode 15.5kHz to 20kHz, the spectral energies of from 11 kHz to 15.5 kHz and from 15.5 kHz to 20 kHz are calculated, and then the ratio of the two energies is coded using 4 bits after being quantized.

5.2.6.2.1.1 Windowing and time-to-frequency transformation

The input high-pass filtered signal is delayed bysamples and windowed to obtain the windowed input signal as shown in figure , where is equal to:

 ()



Figure 51: Windowing of the input high-pass filtered signal

A 640-point length MDCT on top of is used for SWB FD BWE. Refer to subclause 5.3.2.

5.2.6.2.1.2 Transient detection

The input time-domain SHB signal of current frame, sampled at 16kHz, is first high-pass filtered; the high-pass filter serves as a precaution against low frequency components adversely affecting the processing. A first order IIR filter is used, and it is given by:  ()

The output of the high-pass filter  is obtained according to:  ()

where  denotes the 20ms frame length at 16kHz. The high-pass filtered signal is divided into four sub-frames; each corresponding to 5 ms or 80 samples.

The energy of each sub-frame, , is computed according to:

 ()

For each sub-frame, the signal’s long term energy, , is updated according to the following equation:

 ()

In the above equation, the forgetting factor  is set to 0.25, and the convention is that for the first sub-frame,

 from the previous frame. It should be noted that when the current frame and the previous frame apply different BWE algorithms, or the core configurations are different, then the signals’ long term energy is calculated as .

The memory state of the high-pass filter, and are saved for the next frame’s processing. For each sub-frame , a comparison between the short term energy  and the short term energy of previous sub-frame  or the long term energy is performed to detect whether the current frame is TRANSIENT or not. A transient is detected whenever the energy ratio is above a certain threshold which is larger than 1. Formally, a transient is detected whenever:

 ()

where  is the energy ratio threshold and is set to  for INACTIVE frames and otherwise.

It should be noted that if the previous frame did not use SWB Multi-mode FD BWE then the current frame is not classified as a transient frame. In general, the time-frequency transform is applied on a 40ms frame; therefore, a transient affects two consecutive frames. To overcome this, a hangover for a detected transient is applied. A transient detected at a certain frame also triggers a transient in the next frame.

The output of the transient detector is a flag, denoted . The flag is set to the logical value *TRUE* if a transient is detected or *FALSE* otherwise.

In addition, the parameters of spectral tilt and frame class for the current frame are also used for further refinement of the transient decision.

The spectral tilt of the low frequency signal  is calculated by:

 ()

where, and are calculated by:

 ()

is initialized to, and if , is adjusted by adding .

The spectral tilt and class of the current frame are checked against threshold values, and the high frequency TRANSIENT signal classification of the current frame is adjusted. Formally, when is *TRUE*, and  , thesignal classification is set to *FALSE*, andalso the hangover is set to 0.

Another flag, , is used to indicate whether a transient is present in the low frequency signal of the current frame. It is calculated as follows: When the conditionis satisfied, then the flag is set to logical *TRUE;* and it is set to *FALSE* otherwise*.* Here,, and the convention is that for the first sub-frame, from the previous frame.

5.2.6.2.1.3 Frequency domain classification and coding

In frames that are classified as containing transients, the value of  is set (=TRANSIENT). For frames without transients, a frequency sharpness parameter is computed to reflect the spectral fluctuation of the frequency coefficients in the super high band signal and those frames are categorized in one of three classes:

a) HARMONIC: when frequency sharpness is high.

b) NOISE: when frequency sharpness is low.

c) NORMAL: when frequency sharpness is moderate.

The 288 MDCT coefficients in the 6400-13600 Hz frequency range,  are split into nine sharpness bands (32 coefficients per band). The frequency sharpness,, is then defined as the ratio of the peak magnitude to the average magnitude in a sharpness band 

 ()

where the maximum magnitude of spectral coefficients in a sharpness band, denoted, is given by

 ()

Then, six parameters are calculated; the global gain, , the average, the , and three further sharpness parameters are determined; the maximum sharpness,  , the sharpness band counter, , and the noise band counter, .

The maximum sharpness,, in all sharpness bands is computed as:

 ()

The counteris computed from the nine frequency sharpness parameters,, and from the nine maximum magnitudes, , as follows: Initialized to zero,  is incremented by one for each ,, ifand.

The counter is computed from the nine frequency sharpness parametersas follows: Initialized to zero,  is incremented by one for each,if is less than 3.

The class of non-TRANSIENT frames is determined from the three sharpness parameters, , and and four other parameters; the previous frame saved class, , , , and the ratio of the global gains in the current and previous frames.

The threshold of the number of harmonics and the threshold of the maximum sharpness are set according to the signal class of previous frame and the mode of previous extension layer. When the bandwidth is changed i.e., , the  and  will be decreased for harmonic mode and increased for other signal mode:

 ()

 ()

– If, and , the current frame is classified as HARMONIC frame (= HARMONIC) and the counter for signal classis incremented by one whenis less than twelve. Otherwise, is decremented by one whenis larger than zero.

– Then, if, the current frame is also classified as HARMONIC frame (= HARMONIC).

– For other cases, depending on the noise counter,, , and the spectral tilt**,**the current frame is classified as NORMAL or NOISE:

if,and, the current frame is classified as NOISE frame (= NOISE), otherwise, the current frame is classified as NORMAL frame (= NORMAL).

Two bits are transmitted for SHB signal class coding. Table gives the coded bits for each class.

Table 63: SHB signal class coding

|  |  |
| --- | --- |
| Signal class | Coded bits |
| NOISE | 00 |
| TRANSIENT | 01 |
| NORMAL | 10 |
| HARMONIC | 11 |

The signal class  of the current frame is preserved as  for the next frame.

5.2.6.2.1.4 Sub-band division

The 320 MDCT coefficients (at 13.2kbps in the 6150-14150 Hz frequency range,  at 32kbps in the 8000-16000 Hz frequency range) are either split into four sub-bands for TRANSIENT frames or fourteen sub-bands for non-TRANSIENT frames. Table and table define the sub-band boundaries and sizes for TRANSIENT frames and non-TRANSIENT frames respectively. The-th sub-band comprises coefficients  where .

Table : Sub-band boundaries and number of coefficients per sub-band in TRANSIENT frames

|  |  |  |
| --- | --- | --- |
| *j* |  |  |
| 0 | 0 | 76 |
| 1 | 76 | 76 |
| 2 | 152 | 84 |
| 3 | 236 | 84 |
| 4 | 320 | - |

Table 65: Sub-band boundaries and number of coefficients per sub-band in Non-TRANSIENT frames

|  |  |  |
| --- | --- | --- |
| *j* | *)* |  |
| 0 | 0 | 16 |
| 1 | 16 | 24 |
| 2 | 40 | 16 |
| 3 | 56 | 24 |
| 4 | 80 | 16 |
| 5 | 96 | 24 |
| 6 | 120 | 16 |
| 7 | 136 | 24 |
| 8 | 160 | 24 |
| 9 | 184 | 24 |
| 10 | 208 | 24 |
| 11 | 232 | 24 |
| 12 | 256 | 32 |
| 13 | 288 | 32 |
| 14 | 320 | - |

5.2.6.2.1.5 Spectral envelope calculation and quantization

The spectral envelope, or the energy of each band, is computed as follows:

 ()

If the current frame is a Non-TRANSIENT frame, energy control is performed to prevent too much noise being applied to the generated spectrum. The energy control adjusts the spectral energy of each band, depending on the different characteristics between the original high frequency spectrum and the base excitation spectrum.

First a spectral copy is created by mapping the frequencies, depending upon the bandwidth, the ACELP coding modes and the FD BWE mode as defined in table 66.

Table 66: Frequency mapping to generate base excitation spectrum in FD BWE

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mode, Bit-rate, ACELP coding modes | *l* |  |  |  |  |
| WB @ 7.2, 8kbps  All LP-based modes except for AUDIO | 0 | 160 | 239 | 240 | 319 |
| WB @ 7.2, 8kbps  AUDIO  WB @ 13.2kbps  All | 0 | 80 | 159 | 240 | 319 |
| SWB @ 13.2kbps  NORMAL, NOISE | 0 | 112 | 239 | 246 | 373 |
| 1 | 112 | 239 | 374 | 501 |
| 2 | 176 | 239 | 502 | 565 |
| SWB @ 13.2kbps  HARMONIC | 0 | 0 | 239 | 246 | 485 |
| 1 | 128 | 207 | 486 | 565 |
| SWB @ 32kbps  NORMAL, NOISE | 0 | 112 | 239 | 320 | 447 |
| 1 | 112 | 239 | 448 | 575 |
| 2 | 176 | 239 | 576 | 639 |
| SWB @ 32kbps  HARMONIC | 0 | 0 | 239 | 320 | 559 |
| 1 | 128 | 207 | 560 | 639 |

To generate the base excitation spectrum, the spectral copy is normalized by the sum of its absolute spectral components; the window size used depends on the signal characteristics.

 ()

The tonality measures used for the energy control are then calculated:

 ()

The ratio between the tonality of the original high frequency spectrum () and the tonality of the base excitation spectrum () is then calculated as follows:

 (810)

where is 0.35.

The envelope control factor  is then applied to the envelope :

 ()

Next the spectral envelope is adjusted by subtracting the mean vectors which are shown in table 67.

 ()

Table 67: Mean vectors in FD BWE

|  |  |  |
| --- | --- | --- |
| *j* | TRANSIENT | Non-TRANSIENT |
| 0 | 27.23 | 28.62 |
| 1 | 23.81 | 28.96 |
| 2 | 23.87 | 28.05 |
| 3 | 19.51 | 27.97 |
| 4 | - | 26.91 |
| 5 | - | 26.82 |
| 6 | - | 26.35 |
| 7 | - | 25.98 |
| 8 | - | 24.94 |
| 9 | - | 24.03 |
| 10 | - | 22.94 |
| 11 | - | 22.14 |
| 12 | - | 21.23 |
| 13 | - | 20.40 |

If the current frame is a TRANSIENT frame, the following smoothing processes are applied before the envelope quantization.

– If is *TRUE*, the coder type is INACTIVE and the transient hangover is equal to one, the flag is set to 1, and the time envelope is adjusted as follows:

 ()

– Otherwise, the adjustment is as follows:

 ()

If the current frame is a Non-TRANSIENT frame,

 ()

If the current frame is a TRANSIENT frame, the mean squared error (MSE) criterion is used for the search of the VQ, in a Non-TRANSIENT frame, the weighted mean squared error (WMSE) is used. The weighting serves to emphasise the lower frequency bands and is calculated by two methods; one is a deterministic weighting based solely on the frequency, and the other is a weighting that is calculated based upon the envelope. The first frequency weighting is defined in table 68 and the second frequency weighting is calculated as follows:

 ()

Table 68: Frequency weighting 

|  |  |
| --- | --- |
| *j* | Non-TRANSIENT |
| 0 | 1.0 |
| 1 | 0.97826087 |
| 2 | 0.957446809 |
| 3 | 0.9375 |
| 4 | 0.918367347 |
| 5 | 0.9 |
| 6 | 0.882352941 |
| 7 | 0.865384615 |
| 8 | 0.849056604 |
| 9 | 0.833333333 |
| 10 | 0.818181818 |
| 11 | 0.803571429 |
| 12 | 0.789473684 |
| 13 | 0.775862069 |

The SWB spectral envelope is quantized with a multi-stage split VQ using envelope interpolation as in figure 52. In the first stage, two or three candidates’ (, three in TRANSIENT frame, two in Non-TRANSIENT frame) indices are chosen using the error minimization criterion. The set of candidates with the least quantization error, taking into account all quantization steps, is then selected and the selected indices transmitted.



Figure 52: Envelope VQ in a TRANSIENT frame and a Non-TRANSIENT frame

Again during the first stage, values in even positions are selected and quantized using VQ with 5 bits for Non-TRANSIENT frames and 7 bits for TRANSIENT frames.

 (817)

 (818)

In Non-TRANSIENT frames, the candidate indices from the first stage VQ are defined as . The quantization error is calculated and the error is split into  and and quantized using 7 bits and 6 bits respectively, as follows:

 ()

then;

 ()

The candidate indices from the second stage VQ are defined as  and .

The two quantized values are then combined:

 (821)

At odd positions, an interpolation using boundary values is applied for intra-frame prediction and the predicted error is calculated:

 ()

The errors are then split into  and and quantized using 5 bits and 6 bits respectively.

 ()

The candidate indices from the third stage VQ are defined as. In a TRANSIENT frame only 2 stages of quantization are applied. At odd positions, an interpolation using boundary values is applied for intra-frame prediction and the predicted error is calculated and quantized

 (824)

The candidate indices from the second stage VQ are defined as .

The final selected set of indices  for a Non-Transient frame, or  for a Transient frame are then transmitted.

5.2.6.2.1.6 Time envelope calculation and encoding

In case of TRANSIENT frames, ie, the super higher band (SHB) signal of the input in the previous frame is detected as TRANSIENT and the super higher band (SHB) signal of the input in the current frame is detected as NON-TRANSIENT, or the super higher band (SHB) signal of the input in the current frame is detected as TRANSIENT, the time envelope is also calculated. The time envelope, which represents the temporal energy of the SHB signal, is computed as a set of root mean square (RMS) calculations from each 80 samples of time-domain SHB signal. This results in four time envelope coefficients per frame.

 ()

The time envelope is firstly adjusted by the attenuated value which represents the energy attenuation of the WB signal, and the attenuated value is calculated by the original WB signal and local synthesized WB signal:

 ()

 ()

In order to highlight the characteristics of the transient signal, the time envelope of the transient signals is modified. A reference sub-frame is first selected from the sub-frames of the input transient signal, which has the maximal amplitude value of envelope compared with values of the envelopes of the rest sub-frames. Then the time envelope of the reference sub-frame is increased whilst at the same time the envelopes of the sub-frames before and after the reference sub-frame are decreased. Then, the adjusted time envelopes of the SHB signal of the current frame are quantized and coded into the bitstream. In order to get better transient effect, in sub-frames before and after the reference sub-frame, the difference between the decreased time envelope and the maximum time envelope is greater than a preset threshold.

– If the sub-frame index  which is defined asis less than 4, the time envelope is adjusted by:

 ()

where is the sub-frame index with the maximum time envelope . It should be noted when the sub-frame index,and when the condition  is satisfied, the adjustment of time envelope of sub-frame  can be performed.

– For other cases, to obtain the time envelopes of the SHB signal of the current frame used to encode, the adjustment is as follows:

 ()

and

 ()

In addition, when is *TRUE*, the coder type is INACTIVE and the transient hangover is equal to one, the flag is then set to 1, and the time envelope is adjusted as follows

 ()

Finally, the values are further bounded in the range [0,...,15]: ,.

The adjusted time envelopes are rounded and quantized with four bits using uniform scalar quantization in the case of TRANSIENT frames.

5.2.6.2.1.7 Bit allocation for FD BWE

Table illustrates the BWE bit allocation for TRANSIENT and Non-TRANSIENT frames.

Table 69: SWB FD BWE bit allocation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Signal class | Signal class bits () | Time envelope () | Spectral envelope () | Total bits |
| TRANSIENT | 2 | 16 (=4×4) | 13 (=7+6) | 31 |
| NON-TRANSIENT | 2 | 0 | 29 (=5+7+6+5+6) | 31 |

##### 5.2.6.2.2 WB Multi-mode FD Bandwidth Extension

At 13.2kbps, for frame declared as HARMONIC (HM) frame or NORMAL (NM), a bit budget of 6 bits is allocated to the WB Multi-mode FD Bandwidth Extension. One bit is allocated to the signal class and five bits are allocated to encode two spectral envelopes which are calculated by the 80 MDCT coefficients of the higher band (HB) signal, .

At 7.2kbps or 8kbps, it is blind BWE and no bit budget is allocated. In this case, a two-stage blind BWE is used. In the first stage, the high band frequency generation is the same as the BWE in AMR-WB [9], described in subclauses 6.3.1, 6.3.2.2 and 6.3.3 of [9], and it is added to the ACELP core synthesis. Then the second stage BWE is generated as described in the following sub-clauses, and it is added to the core synthesis with the first stage BWE. At 5.9 kbps VBR coding and CNG coding up to 8.0 kbps, the BWE is also blind with no bit budget allocated, but only the first stage BWE is used.

5.2.6.2.2.1 Windowing and time-to-frequency transformation

The input high-pass filtered signal is delayed bysamples and windowed to obtain the windowed input signal as shown in figure , where .

320-point length MDCT on top of  is used for WB FD BWE. Refer to subclause 5.3.2.

5.2.6.2.2.2 Frequency domain classification and coding

At 13.2kbps, frequency domain classification is performed. A frequency sharpness parameter is computed to reflect the spectral fluctuation of the frequency coefficients in the higher band signal and those frames are categorized in one of two classes:

a) HARMONIC: when frequency sharpness is high.

b) NORMAL: when frequency sharpness is moderate.

The 96 MDCT coefficients in the 5600-8000 Hz frequency range  are split into three sharpness bands (32 coefficients per band). The frequency sharpness,, is then defined as the ratio of the peak magnitude to the average magnitude in a sharpness band 

 ()

where the maximum magnitude of spectral coefficients in a sharpness band, denoted, is given by

 ()

Then, another two sharpness parameters are determined: the maximum sharpness,  , and the sharpness band counter, .

The maximum sharpness,, in all sharpness bands is computed as:

 ()

The counteris computed from the three frequency sharpness parameters, , and from the three maximum magnitudes, , as follows: Initialized to zero,  is incremented by one for each  ifand.

The class of HARMONIC frame is determined from these three sharpness parameters, , and the class of the previous frame, .

The threshold of the number of harmonics  and the threshold of the maximum sharpness are set according to the class of the previous frame and the mode of previous extension layer:

 ()

 ()

Initialize the signal class of the current frame as NORMAL frame (= NORMAL).

– If and , the current frame is classified as HARMONIC frame (= HARMONIC) and the counter for signal classis initialized to 0, and incremented by one whenis less than twelve. Otherwise, is decremented by one whenis larger than zero.

– If the counter for signal class is not less than 2, the current frame is also classified as HARMONIC.

One bit is transmitted for HB signal class coding. Table 70 gives the coded bit for each class.

Table 70: HB signal class coding

|  |  |
| --- | --- |
| Signal class *Fclass* | Coded bit |
| NORMAL | 0 |
| HARMONIC | 1 |

The signal class  of the current frame is preserved as  for the next frame.

5.2.6.2.2.3 Spectral envelope calculation and quantization

The spectral envelopes are computed from each 40 samples of frequency-domain HB signal. This results in two spectral envelopes per frame.

 ()

Then envelope control is performed to prevent too much noise being applied to the reconstructed spectrum. First a spectral copy is created by mapping the frequencies, depending upon the bandwidth, the ACELP coding modes and the FD BWE mode as defined in table 66.

To generate the base excitationexcitation spectrum, the spectral copy is normalized by the sum of its absolute spectral components. The parameter of adaptive normalization length  is calculated depending on the original WB MDCT coefficients:

– The 256 WB MDCT coefficients in the 0-6400 Hz frequency range,  are split into 16 sharpness bands (16 coefficients per band). In sharpness band *j*, if  and , the counter is incremented by one.

where, and the maximum magnitude of the spectral coefficients in a sharpness band, denoted, is:

 ()

Parameteris initialized to 0 and calculated for every frame.

– Then the normalization length is obtained:

 ()

where the current normalization length  is calculated depending on the HB signal class:

 ()

and the current normalization length is preserved asfor the next frame.

Then the base excitation spectrum is obtained by:

 ()

where, are the spectral copy coefficients, and the normalized envelope is calculated by:

 ()

The tonality measures used for the energy control are then calculated:

 (843)

The ratio between the tonality of the original high frequency spectrum () and the tonality of the base excitation spectrum () is then calculated as follows:

Void (844)

 (845)

where is 0.35.

The envelope control factor  is then applied to the envelope:

 ()

Then the spectral envelope of log-domain is obtained by:

 ()

Finally, the spectral envelopes are quantized with a 64-dimentional array  described in table 71.

The distance between the spectral envelopes and the codebook is calculated by:

 ()

and the index  is encoded with 5 bits.

Table 71: Codebook 

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 |
|  | 1.1606680 | 0.6594560 | -4.9874350 | -5.1700310 | 10.230799 | -0.0125740 |
|  | 6 | 7 | 8 | 9 | 10 | 11 |
|  | 10.605126 | 9.7910260 | -0.3739880 | -0.6027910 | 6.2753817 | 0.3307670 |
|  | 12 | 13 | 14 | 15 | 16 | 17 |
|  | 9.4537100 | 8.8558020 | 2.9320890 | 2.1643160 | 3.1332030 | 2.9710870 |
|  | 18 | 19 | 20 | 21 | 22 | 23 |
|  | 8.061906 | -0.5905290 | 15.754963 | 5.0496380 | 17.227070 | 18.329395 |
|  | 24 | 25 | 26 | 27 | 28 | 29 |
|  | -2.4710190 | -3.1725330 | -1.4136470 | -1.9457110 | 15.147771 | 14.506490 |
|  | 30 | 31 | 32 | 33 | 34 | 35 |
|  | 11.358370 | 11.714662 | 9.4275510 | -0.1223030 | 7.0970160 | -1.5805260 |
|  | 36 | 37 | 38 | 39 | 40 | 41 |
|  | 12.498663 | 3.1614850 | 10.349261 | 1.5185040 | 5.3809850 | -1.7341900 |
|  | 42 | 43 | 44 | 45 | 46 | 47 |
|  | 1.1224600 | -2.2397020 | 12.362551 | 12.133788 | 4.2788690 | -1.7729040 |
|  | 48 | 49 | 50 | 51 | 52 | 53 |
|  | 6.1577130 | 5.4971410 | 3.3243130 | -2.5710470 | 19.097071 | 9.3576920 |
|  | 54 | 55 | 56 | 57 | 58 | 59 |
|  | 7.6509204 | 7.4404626 | 0.5055090 | -3.7073090 | 18.584702 | 11.302494 |
|  | 60 | 61 | 62 | 63 |  |  |
|  | 18.706564 | 18.308905 | 23.010420 | 22.915377 |  |  |

5.2.6.2.2.4 Bit allocation for FD BWE

Table 72 illustrates the WB BWE bit allocation.

Table 72: WB FD BWE bit allocation

|  |  |  |  |
| --- | --- | --- | --- |
| Signal class | Signal class bit () | Spectral envelope () | Total bits |
| HARMONIC | 1 | 5 | 6 |
| NORMAL | 1 | 5 | 6 |

#### 5.2.6.3 Coding of upper band at 64 kb/s

The SWB, resp. FB, signal at 64 kbps bit-rate is coded in two bands. The lower band that covers 0-8kHz is coded using the LP-based coding at 16 kHz internal sampling rate as described earlier and the upper band that extends the coded band-width up to 16 kHz, resp. 20 kHz, is coded using a high-rate upper band coding. The same upper band coding with a fixed 16 kbps bit-budget is used in all GC, TC and IC frames. This bit-budget can be eventually increased with unused bits coming from the AVQ within the combined algebraic codebook.

The upper band coding is mostly done in the MDCT domain and has two modes: normal and transient. While normal mode is used in most of generic and voiced frames, the use of transient mode minimalizes the pre-echo and post-echo in frames where the signal at the frame beginning is significantly different from the signal at the frame end, e.g. onsets. Detection of transient frames is done in time domain using a detector described in .

First the input signal, filtered by the HP filter and sampled at 32 or 48 kHz,  is transformed using MDCT and OLA function. In normal mode, the whole frame is transformed at once while in transient mode the frame is divided into four sub-frames and thus four sets of spectral coefficients  are present. In both modes only spectrum coefficients between 7.6 kHz and 14.4 kHz are encoded. While the spectrum between 7.6 kHz and 14.4 kHz is divided in normal mode into four bands of 1.7 kHz width each, it is divided into two bands of 3.4 kHz each in transient mode. The other frequency coefficients are zeroed. Consequently spectral coefficients, , are encoded in normal mode frames and spectral coefficients , , are encoded in transient mode frames.

##### 5.2.6.3.1 Coding in normal mode

In normal mode frames, the global gain is computed on the spectrum 7.6 kHz – 14.4 kHz as follows

 (849)

and quantized using a 5-bit log gain quantizer at the range of [3.0; 500.0].

The quantized global gain  is further used to normalize the spectrum  resulting in a normalized spectrum by the quantized global gain.

 ()

where  is the start frequency bin of spectrum reconstruction and  for normal frames.

Then the spectrum envelope is computed in four bands which results in 68 spectral coefficients per band.

 ()

The spectrum envelope is quantized using two two-dimensional VQs by means of 6 bits codebook and 5 bits codebook in table 73 and table 74, respectively

Table 73: 6 bits spectral envelope VQ codebook

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | |  |  | |  |  | |
| 0 | 0.044983 | 0.0417 | 22 | 8.919388 | 9.762914 | 44 | 15.26931 | 21.53914 |
| 1 | 0.524276 | 0.469365 | 23 | 11.29932 | 11.7639 | 45 | 16.98352 | 24.69959 |
| 2 | 0.671757 | 0.605513 | 24 | 11.78222 | 5.879754 | 46 | 19.59173 | 22.68968 |
| 3 | 0.983501 | 0.855093 | 25 | 14.05046 | 9.665228 | 47 | 20.1462 | 25.88847 |
| 4 | 1.227874 | 1.1322 | 26 | 11.20153 | 9.001128 | 48 | 17.79742 | 19.45312 |
| 5 | 1.672212 | 1.432704 | 27 | 14.43475 | 13.23657 | 49 | 21.29062 | 20.18658 |
| 6 | 2.548211 | 2.361091 | 28 | 14.33726 | 3.904411 | 50 | 24.09732 | 19.08672 |
| 7 | 3.196961 | 3.306999 | 29 | 20.07105 | 4.335061 | 51 | 23.61309 | 22.54586 |
| 8 | 2.580753 | 5.217478 | 30 | 18.10581 | 8.223599 | 52 | 23.68201 | 16.32824 |
| 9 | 4.207751 | 7.243802 | 31 | 22.35229 | 9.603263 | 53 | 26.88655 | 19.40244 |
| 10 | 3.517157 | 1.738487 | 32 | 7.242756 | 16.56449 | 54 | 26.00977 | 15.63221 |
| 11 | 4.381567 | 2.753657 | 33 | 11.77753 | 19.16765 | 55 | 28.93993 | 16.24062 |
| 12 | 4.758266 | 4.696094 | 34 | 11.1218 | 15.45598 | 56 | 25.09448 | 12.36642 |
| 13 | 6.827988 | 6.106459 | 35 | 14.56358 | 17.35957 | 57 | 27.71338 | 13.26328 |
| 14 | 4.450459 | 10.13121 | 36 | 17.82122 | 11.89472 | 58 | 28.33095 | 10.32926 |
| 15 | 7.256045 | 12.48804 | 37 | 17.46603 | 15.29606 | 59 | 30.63283 | 12.85128 |
| 16 | 6.70872 | 1.953339 | 38 | 21.33696 | 13.45518 | 60 | 25.2738 | 6.138124 |
| 17 | 6.60403 | 3.69956 | 39 | 20.54434 | 17.12537 | 61 | 29.19534 | 7.222413 |
| 18 | 10.61273 | 2.537916 | 40 | 9.056358 | 22.33831 | 62 | 32.17132 | 5.019567 |
| 19 | 9.387467 | 4.241173 | 41 | 11.23842 | 28.83252 | 63 | 31.979 | 9.473855 |
| 20 | 7.119045 | 8.281485 | 42 | 13.26273 | 25.14338 |  |  |  |
| 21 | 9.062854 | 7.086526 | 43 | 16.24356 | 28.25685 |  |  |  |

Table 74: 5 bits spectral envelope VQ codebook

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | |  |  | |
| 0 | 0.512539 | 0.472507 | 16 | 13.67263 | 5.457414 |
| 1 | 1.338963 | 1.108591 | 17 | 16.47199 | 3.917684 |
| 2 | 2.544041 | 1.759765 | 18 | 20.91033 | 6.43281 |
| 3 | 3.124053 | 3.045299 | 19 | 25.45733 | 8.61722 |
| 4 | 4.892713 | 3.721097 | 20 | 16.4107 | 7.574456 |
| 5 | 4.010297 | 5.750862 | 21 | 18.57439 | 10.2915 |
| 6 | 5.111215 | 2.164709 | 22 | 22.08876 | 12.51216 |
| 7 | 6.667518 | 3.893404 | 23 | 21.17053 | 17.20871 |
| 8 | 8.454117 | 2.75143 | 24 | 5.276107 | 9.62247 |
| 9 | 11.12357 | 3.518174 | 25 | 9.093585 | 11.27469 |
| 10 | 6.622948 | 5.960704 | 26 | 11.94566 | 15.53814 |
| 11 | 8.562429 | 5.003579 | 27 | 16.55041 | 15.04656 |
| 12 | 8.919363 | 7.784057 | 28 | 6.358148 | 17.5474 |
| 13 | 10.75904 | 5.959438 | 29 | 13.31662 | 21.76552 |
| 14 | 12.44919 | 8.359519 | 30 | 7.646096 | 26.10672 |
| 15 | 13.67701 | 11.23058 | 31 | 2.451297 | 31.9331 |

The quantized spectrum envelope is used to further normalize the spectrum resulting in spectrum normalized per bands.

 ()

Afterwards, the number of the bands to be quantized is calculated according to the total bits and the saturated threshold, and the bands are selected according to the quantized spectrum envelopes. Once the bands are selected, the first stage encoding is processed by means of AVQ. If there are at least 14 remaining bits after the first stage encoding and the first stage quantized spectrum is non-zero, a second stage encoding is employed also by means of AVQ.

Calculate the number of the bands to be quantized according to total bits  and the saturated threshold, and select the bands to be quantized according to the quantized envelopes as follows:

– If , the selected bands are by:

 ()

– Otherwise, the selected bands are by:

 ()

where the and are set as follows:

 ()

 ()

where  is the sub-band index with the minimum envelope , and it is calculated by:

 ()

The number of the sub-bands is obtained according to the number of the total bits  and the saturated threshold  as follows:

 ()

Quantize the normalized coefficients of the selected bands (sub-bands) by AVQ to obtain the quantized normalized coefficients.

The envelope of the spectrum between 14.4 kHz and 16 kHz in SWB is predicted by:

 ()

And the envelope of the spectrum between 14.4 kHz and 20 kHz in FB is calculated and quantized as follows:

 ()

where  is the start frequency bin of spectrum reconstruction and  for normal frames.  is the width of the MDCT coefficients between 14.4 kHz and 20 kHz in FB, and is set by:.

The ratio of the envelopes is refined:

 ()

And the index of attenuation factor  is obtained according to the ratio of the envelopes, and the envelope of the spectrum between 14.4 kHz and 20 kHz in FB is finally obtained according to the attenuation factor:

 ()

 ()

Then the index is encoded with 2bits.

If the number of the remaining bits after the first stage encoding is larger than 14, then the second stage encoding is needed. Select  sub-bands from the first stage selected sub-bands to perform the second stage encoding, and the number of the sub-bands is calculated according to the number of remaining bitsand the saturated threshold .

The input coefficients  of the second stage encoding are obtained by reordering the differences between the original normalized coefficients  and the quantized normalized coefficients  as follows:

First, in the sub-bands with the AVQ codebook index,

 ()

where the index is initialized to 0, and is incremented by 1 when.

Then, in the sub-bands , if , , and the index  is incremented by 1.

The number of the sub-bands  is calculated according to the number of the remaining bits  and the saturated threshold:

 ()

The coefficients of the first  sub-bands in  are selected to perform the second stage encoding as follows:

The global gain of second stage encodingis calculated and quantized as.

 ()

and the spectrum is normalized with as follows:

 ()

Then the normalized spectrum is quantized by AVQ.

##### 5.2.6.3.2 Coding in transient mode

In transient mode frames, a similar procedure as in normal mode frames is employed and the following descriptions focus on the differences.

The total bit-budget is divided by four in order to obtain a bit-budget for every sub-frame and the encoding is performed four times (once for every sub-frame). In sub-frame,,the global gain is computed on the spectrum 7.6 kHz – 14.4 kHz using equation () for and quantized using a 5-bit log gain quantizer at the range of [3.0, 500.0].

 ()

where  is the start frequency bin of spectrum reconstruction and  for transient frames. is the frame length. for SWB signal and for FB signal

The spectrum envelope is computed in two bands for each sub-frame, thus each band contains 34 spectral coefficients.

Then, the spectral envelope of normalized spectrum is calculated:

 ()

where,

 ()

The total 8 spectral envelopes in 4 sub-frame are divided into 4 two-dimensional vectors, i.e., , ,  in each sub-frame are combined as a vector, and quantized using two-dimensional VQs. For the first vector, the spectral envelopes in the first sub-frame  are quantized using one two-dimensional VQ by means of 4 bits codebook defined in table 75.

Table 75: 4 bit spectral envelope VQ codebook

|  |  |  |
| --- | --- | --- |
|  |  | |
| 0 | 0.799219 | 0.677609 |
| 1 | 1.754571 | 1.215689 |
| 2 | 2.846222 | 2.017775 |
| 3 | 4.379336 | 1.975914 |
| 4 | 5.935472 | 2.945818 |
| 5 | 3.938621 | 4.220399 |
| 6 | 8.080808 | 2.632276 |
| 7 | 7.579771 | 4.986835 |
| 8 | 4.956485 | 10.36366 |
| 9 | 7.739148 | 8.652471 |
| 10 | 9.238397 | 7.051655 |
| 11 | 10.205707 | 5.619638 |
| 12 | 10.645117 | 4.374648 |
| 13 | 11.66018 | 3.474015 |
| 14 | 10.845836 | 2.664596 |
| 15 | 11.724073 | 1.637023 |

The index of this VQs is noted as . This 4 bits codebook can be divided into two 3 bits codebook. The first 3 bits codebook is , and the second is  in the table 75. Then the first quantized vector is determined one of the 3 bits codebook according to the .

If , the first 3 bits codebook is selected as new codebook;

if , the second 3 bits codebook is selected as new codebook.

Then the following three vectors are quantized by using the 3 bits codebook determined before.

The quantized spectral envelope is applied to the spectrum:

 ()

where ,  and .

Then, the normalized spectrum is quantized by AVQ.

If , the envelope of the spectrum between 14.4 kHz and 16 kHz in SWB is predicted by:

 ()

If , the envelope of the spectrum between 14.4 kHz and 20 kHz in FB is calculated and quantized as follows:

 ()

where,  is the start frequency bin of spectrum reconstruction and  for transient frames.  is the width of the MDCT coefficients between 14.4 kHz and 20 kHz in FB, and is set by.

The ratio of the envelopes is refined as:

 ()

And the index of attenuation factor  is obtained according to the ratio of the envelopes, and the envelope of the spectrum between 14.4 kHz and 20 kHz in FB is finally obtained according to the attenuation factor :

 ()

 ()

where . Then the index is encoded with 2bits.

Finally, the unused AVQ bits from the current sub-frame are employed in the subsequent sub-frame within the same frame.