## 5.7 AMR-WB-interoperable modes

The EVS codec supports modes to allow for interoperability with AMR-WB (and ITU-T G.722.2). The inclusion of the interoperable modes has been streamlined due to the fact that the core ACELP coder described in subclause 5.2 is similar to AMR-WB (when operating at 12.8 kHz internal sampling, using the same pre emphasis and perceptual weighting, etc.).

### 5.7.1 Pre-processing

The high-pass filtering, sampling conversion, pre-emphasis, spectral analysis, signal activity detection functions are the same as those described in subclause 5.1.

### 5.7.2 Linear prediction analysis and quantization

#### 5.7.2.1 Windowing and auto**‑**correlation computation

Short-term prediction analysis is performed once per speech frame using the autocorrelation approach per subclause 5.1.9. However, the 30 ms asymmetric window defined in subclause 5.2.1 of [9] and a look‑ahead of 5 ms are used in the autocorrelation calculation. The frame structure is depicted in figure 1.



Figure 86: Relative positions and length of the LP analysis windows for the AMR-WB interoperable option

The autocorrelations of the windowed signal are computed in the same way as described in subclause 5.1.9.2, except that = 384 in equation (45). Note that the autocorrelations are computed in the same way as in subclause 5.2.1 of [9] but with a different white noise correction factor value and lag windowing as described in subclause 5.1.9.3 of this Specification.

#### 5.7.2.2 Levinson**‑**Durbin algorithm

The Levinson-Durbin algorithm is the same as in subclause 5.5.1.3.

#### 5.7.2.3 LP to ISP conversion

For a linear predictive model *A(z)* of order *m* we can define the line spectral polynomials as

(1405)

where *l=1* for the line spectrum polynomials and *l=0* for the immittance spectrum polynomials. These polynomials *P(z)* and  *Q(z)* are symmetric and anti-symmetric, respectively, such that the point of symmetry is . It follows that when evaluating at the unit circle then the obtained spectra of and will be real and imaginary, respectively. Further, since polynomials *P(z)* and  *Q(z)* have roots on the unit circle, they can be located by a zero-crossing search of the two spectra.

The evaluation of on the unit circle can be implemented with an FFT of length *N=256*. Since the two spectra are imaginary, we can evaluate both spectra simultaneously, that is, since , we can determine the spectrum of 2 by an FFT, multiply by and then obtain the two spectra in the real and imaginary parts. Scaling by the factor 2 does not influence location of zeros, whereby it can be omitted.

To reduce numerical range of the spectrum, we can convolve by a filter , where

(1406)

and the constants are and . That is, we calculate the spectrum of and multiply with the phase-shift .

When calculating the FFT, we can reduce complexity by applying pruning methods. That is, since is a sequence of length , but the FFT is of length , we can omit all those operations which involve computations with the  zeros.

#### 5.7.2.4 ISP to LP conversion

The ISP to LP conversion is the same as in subclause 5.2.4 of [9].

#### 5.7.2.5 Quantization of the ISP coefficients

For interoperability reasons, ISF quantization is the same as in subclause 5.2.5 of [9].

#### 5.7.2.6 Interpolation of the ISPs

The set of LP parameters is used for the 4th subframe, whereas the 1st, 2nd and 3rd subframes use a linear interpolation of the parameters in the adjacent frames. The interpolation is performed on the ISPs in the domain. Let be the ISP vector at the 4th subframe of the current frame, and is the ISP vector at the 4th subframe of the previous frame. The interpolated ISP vectors at the 1st, 2nd and 3rd subframes are given by

(1407)

The same formula is used for interpolation of both quantized and unquantized ISPs. The interpolated ISP vectors are used to compute a different LP filter at each subframe (both quantized and unquantized) using the ISP to LP conversion method described in subclause 5.2.4 of [9].

### 5.7.3 Perceptual weighting

Perceptual weighting is performed as described in subclause 5.1.10.1 for a sub-frame size = 64.

### 5.7.4 Open-loop pitch analysis

The open-loop pitch analysis is performed as described in subclause 5.1.10.

### 5.7.5 Impulse response computation

Same as subclause 5.2.3.1.3.

### 5.7.6 Target signal computation

Same as subclause 5.2.3.1.2.

### 5.7.7 Adaptive codebook search

Same as subclause 5.2.3.1.4.

### 5.7.8 Algebraic codebook search

For interoperability reasons, the algebraic codebook structure and pulse indexing is the same as clauses 5.8.1 and 5.8.2 of [9]. The algebraic codebook search procedure is the same as described in clause 5.8.3 of [9] except the pulse-sign pre-selection described in the last paragraph of Clause 5.2.3.1.5.9 (The search criterion at lower bitrates), which is also used at the lowest bit-rate of AMR-WB-interoperable modes.

### 5.7.9 Quantization of the adaptive and fixed codebook gains

For interoperability reasons, the quantization of gains is conducted in the same manner as described in subclause 5.9 of [9].

### 5.7.10 Memory update

The memory update for AMR-WB interoperable modes is similar to subclause 5.10 of [9], however some extra states defined in the EVS codec are also updated to maintain a consistent operation for the next coding frame when the last frame was coded with an AMR-WB interoperable mode.

### 5.7.11 High-band gain generation

For interoperability reasons, the quantization of high-band gain in each 5ms sub-frame is conducted in the same manner as described in subclause 5.11 of [9].

### 5.7.12 CNG coding

The CNG encoding in AMR-WB-interoperable mode is described by referring to subclause 5.6.2 with several differences described below. Instead of the LSF vector which is quantized and transmitted in the SID frame in primary mode, the ISF vector is used for quantization and transmitted in the SID frame in AMR-WB-interoperable mode. 28 bits are used for ISF quantization that is one bit less than the primary mode. The quantization of the ISF vector is described in [9]. The excitation energy used for quantization and transmission in the SID frame is calculated in the same way as described in subclause 5.6.2.1.5. However, the excitation energy is quantized in the AMR-WB-interoperable mode using different numbers of bits and a different quantization step-size from the primary mode. 6 bits are used for quantization in AMR-WB-interoperable mode, instead of the 7 bits used in the primary mode. The quantization step Δ described in subclause 5.6.2.1.5 is 2.625 in the AMR-WB-interoperable mode, instead of 5.25 in primary mode. The quantization index is also limited to [0, 63] in AMR-WB-interoperable mode, instead of to [0, 127] in primary mode. The smoothed LP synthesis filter, , used for local CNG synthesis is obtained in the same way as described in subclause 5.6.2.1.4. And the smoothed quantized excitation energy,, used for local CNG synthesis is also obtained in the same way as described in subclause 5.6.2.1.6. However, the CNG type flag bit, the bandwidth indicator bit, the core sampling rate indicator bit, the hangover frame counter bits and the Low-band excitation spectral envelope bits used in primary SID mode are not encoded into the SID frame in AMR-WB-interoperable mode, but one dithering bit (always set to 0) is always encoded. The high-band analysis and quantization is also ignored for AMR-WB-interoperable mode, so the high-band energy bits are also not encoded into the AMR-WB-interoperable mode SID frame. Local CNG synthesis is performed by filtering the excitation signal (see subclause 6.8.4) through the smoothed LP synthesis filter.

## 5.8 Channel Aware Coding

### 5.8.1 Introduction

EVS offers partial redundancy based error robust channel aware mode at 13.2 kbps for both wideband and super-wideband audio bandwidths. Depending on the criticality of the frame, the partial redundancy is dynamically enabled or disabled for a particular frame, while keeping a fixed bit budget of 13.2 kbps.

### 5.8.2 Principles of Channel Aware Coding

In a VoIP system, packets arrive at the decoder with random jitters in their arrival time. Packets may also arrive out of order at the decoder. Since the decoder expects to be fed a speech packet every 20 msec to output speech samples in periodic blocks, a de-jitter buffer [7] is required to absorb the jitter in the packet arrival time. Larger the size of the de-jitter buffer, the better is its ability to absorb the jitter in the arrival time and consequently, fewer late arriving packets are discarded. Voice communications is also a delay critical system and therefore it becomes essential to keep the end to end delay as low as possible so that a two way conversation can be sustained.

The design of an adaptive de-jitter buffer reflects the above mentioned trade-offs. While attempting to minimize packet losses, the jitter buffer management algorithm in the decoder also keeps track of the delay in packet delivery as a result of the buffering. The jitter buffer management algorithm suitably adjusts the depth of the de-jitter buffer in order to achieve the trade-off between delay and late losses.

EVS channel aware mode uses partial redundant copies of current frames along with a future frame for error concealment. The partial redundancy technology transmits partial copies of the current frame along with a future frame with the hope that in the event of the loss of the current frame (either due to network loss or late arrival) the partial copy from the future frame can be retrieved from the jitter buffer to improve the recovery from the loss.

The difference in time units between the transmit time of the primary copy of a frame and the transmit time of the redundant copy of the frame (piggy backed onto a future frame) is called the FEC offset. If the depth of the jitter buffer at any given time is at least equal to the FEC offset, then it is quite likely that the future frame is available in the de-jitter buffer at the current time instance. The FEC offset is a configurable parameter at the encoder which can be dynamically adjusted depending on the network conditions. The concept of partial redundancy in EVS with FEC offset equal to 3 is shown in figure 87.



Figure 87 : Concept of partial redundancy in channel aware mode

The redundant copy is only a partial copy that includes just a subset of parameters that are most critical for decoding or arresting error propagation.

The EVS channel aware mode transmits redundancy in-band as part of the codec payload as opposed to transmitting redundancy at the transport layer (e.g., by including multiple packets in a single RTP payload). Including the redundancy in-band allows the transmission of redundancy to be either channel controlled (e.g., to combat network congestion) or source controlled. In the latter case, the encoder can use properties of the input source signal to determine which frames are most critical for high quality reconstruction at the decoder and selectively transmit redundancy for those frames only. Another advantage of in-band redundancy is that source control can be used to determine which frames of input can best be coded at a reduced frame rate in order to accommodate the attachment of redundancy without altering the total packet size. In this way, the channel aware mode includes redundancy in a constant-bit-rate channel (13.2 kbps).

### 5.8.3 Bit-Rate Allocation for Primary and Partial Redundant Frame Coding

#### 5.8.3.1 Primary frame bit-rate reduction

A measure of compressibility of the primary frame is used to determine which frames can best be coded at a reduced frame rate. For TCX frame the 9.6kpbs setup is applied for WB as well as for SWB. For ACELP the following apply. The coding mode decision coming from the signal classification algorithm is first checked. Speech frames classified for Unvoiced Coding (UC) or Voiced Coding (VC) are suitable for compression. For Generic Coding (GC) mode, the correlation (at pitch lag) between adjacent sub-frames within the frame is used to determine compressibility. Primary frame coding of upper band signal (i.e., from 6.4 to 14.4 kHz in SWB and 6.4 to 8 kHz in WB) in channel aware mode uses time-domain bandwidth extension (TBE) as described in subclause 5.2.6.1. For SWB TBE in channel aware mode, a scaled down version of the non-channel aware mode framework is used to obtain a reduction of bits used for the primary frame. The LSF quantization is performed using an 8-bit vector quantization in channel aware mode while a 21-bit scalar quantization based approach is used in non-channel aware mode. The SWB TBE primary frame gain parameters in channel aware mode are encoded similar to that of non-channel aware mode at 13.2 kbps, i.e., 8 bits for gain parameters. The WB TBE in channel aware mode uses similar encoding as used in 9.6 kbps WB TBE of non-channel aware mode, i.e., 2 bits for LSF and 4 bits for gain parameters.

#### 5.8.3.2 Partial Redundant Frame Coding

The size of the partial redundant frame is variable and depends on the characteristics of the input signal. Also criticality measure is an important metric. A frame is considered as critical to protect when loss of the frame would cause significant impact to the speech quality at the receiver. The criticality also depends on if the previous frames were lost or not. For example, a frame may go from being non-critical to critical if the previous frames were also lost. Parameters computed from the primary copy coding such as coder type classification information, subframe pitch lag, factor M etc are used to measure the criticality of a frame. The threshold, to determine whether a particular frame is critical or not, is a configurable parameter at the encoder which can be dynamically adjusted depending on the network conditions. For example, under high FER conditions it may be desirable to adjust the threshold to classify more frames as critical. Partial frame coding of upper band signal relies on coarse encoding of gain parameters and interpolation/extrapolation of LSF parameters from primary frame. The TBE gain parameters estimated during the primary frame encoding of the (n – FEC offset )-th frame is re-transmitted during the n-th frame as partial copy information. Depending on the partial frame coding mode, i.e., GENERIC or VOICED or UNVOICED, the re-transmission of the gain frame, uses different quantization resolution and gain smoothing.

The following sections describe the different partial redundant frame types and their composition.

##### 5.8.3.2.1 Construction of partial redundant frame for Generic and Voiced Coding modes

In the coding of the redundant version of the frame, a factor is determined based on the adaptive and fixed codebook energy.

(1408)

In this equation, denotes the adaptive codebook energy and denotes the fixed codebook energy. A low value of indicates that most of the information in the current frame is carried by the fixed codebook contribution. In such cases, the partial redundant copy (RF\_NOPRED) is constructed using one or more fixed codebook parameters only (FCB pulses and gain). A high value of M indicates that most of the information in the current frame is carried by the adaptive codebook contribution. In such cases, the partial redundant copy (RF\_ALLPRED) is constructed using one or more adaptive codebook parameters only (pitch lag and gain). If M takes mid values then a mixed coding mode is selected where one or more adaptive codebook parameters and one or more fixed codebook parameters are coded (RF\_GENPRED). Under Generic and Voiced Coding modes, the TBE gain frame values are typically low and demonstrate less variance. Hence a coarse TBE gain frame quantization with gain smoothing is used.

##### 5.8.3.2.2 Construction of partial redundant frame for Unvoiced Coding mode

The low bit-rate Noise Excited Linear Prediction coding scheme (subclause 5.2.5.4) is used to construct a partial redundant copy for an unvoiced frame type (RF\_NELP). In Unvoiced coding mode, the TBE gain frame has a wider dynamic range. To preserve this dynamic range, the TBE gain frame quantization in Unvoiced coding mode uses a similar quantization range as that of the one used in the primary frame.

##### 5.8.3.2.3 Construction of partial redundant frame for TCX frame

In case of TCX partial redundant frame type, a partial copy consisting of some helper parameters is used to enhance the frame loss concealment algorithm. There are three different partial copy modes available, which are RF\_TCXFD, RF\_TCXTD1 and RF\_TCX\_TD2. Similar to the PLC mode decision on the decoder side, the selection of the partial copy mode for TCX is based on various parameters such as the mode of the last two frames, the frame class, LTP pitch and gain.

5.8.3.2.3.1 Frequency domain concealment (RF\_TCXFD) partial redundant frame type

29 bits are used for the RF\_TCXFD partial copy mode.

• 13bits are used for the LSF quantizer which is the same as used for regular low rate TCX coding.

• The global TCX gain is quantized using 7 bits.

• The classifier info is coded on 2 bits.

5.8.3.2.3.2 Time domain concealment (RF\_TCXTD1 and RF\_TCXTD2) partial redundant frame type

The partial copy mode RF\_TCXTD1 is selected if the frame contains a transient or if the global gain of the frame is much lower than the global gain of the previous frame. Otherwise RF\_TCXTD2 is chosen.

Overall 18bits of side data are used for both modes.

• 9bits are used to signal the TCX LTP lag

• 2 bits for signalling the classifier info

##### 5.8.3.2.4 RF\_NO\_DATA partial redundant frame type

This is used to signal a configuration where the partial redundant copy is not sent and all bits are used towards primary frame coding.

The primary frame bit-rate reduction and partial redundant frame coding mechanisms together determine the bit-rate allocation between the primary and redundant frames to be included within a 13.2 kbps payload.

#### 5.8.3.3 Decoding

At the receiver, the de-jitter buffer provides a partial redundant copy of the current lost frame if it is available in any of the future frames. If present, the partial redundant information is used to synthesize the lost frame. In the decoding, the partial redundant frame type is identified and decoding performed based on whether only one or more adaptive codebook parameters, only one or more fixed codebook parameters, or one or more adaptive codebook parameters and one or more fixed codebook parameters, TCX frame loss concealment helper parameters, or Noise Excited Linear Prediction parameters are coded. If either the current frame or the previous frame adjacent to the current frame is a partial redundant frame, then the decoding parameter of the current frame, such as the LSP parameters, the gain of the adaptive codebook, the gain of the fixed codebook or the BWE gain, is firstly obtained and then post-processed according to the decoding parameters or signal type from previous frames relative to the current frame position and the decoding parameters or signal type from the current frame. Additionally, the decoding parameters or signal type from the future frames relative to the current frame position may also be used for the post-processing of the gains of the adaptive codebook and the fixed codebook, and the spectral tilt may also be used for the post-processing of the gain of the adaptive codebook. The post-processed parameters are used to reconstruct the output signal. Finally, the frame is reconstructed based on the coding scheme. The TCX partial info is decoded, but in contrast to ACELP partial copy mode, the decoder is run in concealment mode. The difference to regular concealment is just that the parameters available from the bitstream are directly used and not derived by concealment.

### 5.8.4 Channel aware mode encoder configurable parameters

The channel aware mode encoder may use the following configurable parameters to adapt its operation to track the channel characteristics seen at the receiver. These parameters maybe computed at the receiver (as described in subclause 6.3) and communicated to the encoder via a receiver triggered feedback mechanism.

Partial redundancy offset (): The difference in time units between the transmit time of the primary copy of a frame and the transmit time of the redundant copy of that frame which is piggy backed onto a future frame is called the partial redundancy offset or FEC offset . The optimal partial redundancy offset that maximizes the probability of partial redundant copy retrival may be computed in the decoder JBM solution and fed back to the encoder as described above.

Frame erasure rate indicator () having the following values: (low) for FER rates <5% or (high) for FER>5%. This parameter controls the threshold used to determine whether a particular frame is critical or not as described in subclause 5.8.3.2. Such an adjustment of the criticality threshold is used to control the frequency of partial copy transmission. The setting adjusts the criticality threshold to classify more frames as critical to transmit as compared to the setting.

It is noted that these encoder configurable parameters are optional with default set to = and =3.