3GPP TS 26.402 V16.0.0 (2020-07)

Technical Specification

3rd Generation Partnership Project;

Technical Specification Group Services and System Aspects;

General Audio Codec audio processing functions;

Enhanced aacPlus general audio codec;

Additional decoder tools

(Release 16)

** 

The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP.   
The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented.   
This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification.  
Specifications and reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices.

Keywords

UMTS, codec, LTE

***3GPP***

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis

Valbonne - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

http://www.3gpp.org

***Copyright Notification***

No part may be reproduced except as authorized by written permission.  
The copyright and the foregoing restriction extend to reproduction in all media.

© 2020, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).

All rights reserved.

UMTS™ is a Trade Mark of ETSI registered for the benefit of its members

3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  
LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners

GSM® and the GSM logo are registered and owned by the GSM Association

Contents

Foreword [4](#__RefHeading___Toc517363334)

1 Scope [5](#__RefHeading___Toc517363335)

2 Normative references [5](#__RefHeading___Toc517363336)

3 Definitions, symbols and abbreviations [5](#__RefHeading___Toc517363337)

3.1 Definitions [5](#__RefHeading___Toc517363338)

3.2 Symbols [6](#__RefHeading___Toc517363339)

3.3 Abbreviations [6](#__RefHeading___Toc517363340)

4 Outline description [7](#__RefHeading___Toc517363341)

5 Error concealment [7](#__RefHeading___Toc517363342)

5.1 AAC error concealment [7](#__RefHeading___Toc517363343)

5.2 SBR error concealment [8](#__RefHeading___Toc517363344)

6 SBR stereo parameter to mono parameter downmix [11](#__RefHeading___Toc517363345)

6.1 Inverse filtering [11](#__RefHeading___Toc517363346)

6.2 Additional harmonics [11](#__RefHeading___Toc517363347)

6.3 Envelope time borders [12](#__RefHeading___Toc517363348)

6.4 Noise time borders [12](#__RefHeading___Toc517363349)

6.5 Envelope data [13](#__RefHeading___Toc517363350)

6.6 Noise floor data [14](#__RefHeading___Toc517363351)

7 Output resampler tool [14](#__RefHeading___Toc517363352)

7.1 QMF bandlimiter [14](#__RefHeading___Toc517363353)

7.2 Spline resampler [15](#__RefHeading___Toc517363354)

7.3 Postfilter [15](#__RefHeading___Toc517363355)

Annex A (informative): Change history [17](#__RefHeading___Toc517363356)

# Foreword

The present document describes tools used in the Enhanced aacPlus general audio codec for the general audio service within the 3GPP system.

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of this TS, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 Indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the specification;

# 1 Scope

This Telecommunication Standard (TS) describes the error concealment algorithm, SBR parameter downmix and output resampling for the Enhanced aacPlus general audio codec [3].

# 2 Normative references

This TS incorporates by dated and undated reference, provisions from other publications. These normative references are cited in the appropriate places in the text and the publications are listed hereafter. For dated references, subsequent amendments to or revisions of any of these publications apply to this TS only when incorporated in it by amendment or revision. For undated references, the latest edition of the publication referred to applies.

[1] ISO/IEC 14496-3:2001/Amd.1:2003: "Bandwidth Extension".

[2] ISO/IEC 14496-3:2001/Amd.1:2003/DCOR1.

[3] 3GPP TS 26.401: "Enhanced aacPlus general audio codec; General Description".

# 3 Definitions, symbols and abbreviations

## 3.1 Definitions

For the purposes of this TS, the following definitions apply:

**band:** (as in limiter band, noise floor band, etc.) a group of consecutive QMF subbands

**envelope scalefactor:** an element representing the averaged energy of a signal over a region described by a frequency band and a time segment

**frequency band:** interval in frequency, group of consecutive QMF subbands

**frequency border:** frequency band delimiter, expressed as a specific QMF subband

**noise floor:** a vector of noise floor scalefactors

**noise floor scalefactor:** an element associated with a region described by a frequency band and a time segment, representing the ratio between the energy of the noise to be added to the envelope adjusted HF generated signal and the energy of the same

**SBR envelope:** a vector of envelope scalefactors

**SBR frame:** time segment associated with one SBR extension data element

**SBR range:** the frequency range of the signal generated by the SBR algorithm

**subband:** a frequency range represented by one row in a QMF matrix, carrying a subsampled signal

**time border:** time segment delimiter, expressed as a specific time slot

**time segment:** interval in time, group of consecutive time slots

**time / frequency grid:** a description of SBR envelope time segments and associated frequency resolution tables as well as description of noise floor time segments

**time slot:** finest resolution in time for SBR envelopes and noise floors. One time slot equals two subsamples in the QMF domain

## 3.2 Symbols

For the purposes of this TS, the following symbols apply:

**E***Orig* has *LE* columns where each column is of length *NLow* or *NHigh* depending on the frequency resolution for each SBR envelope. The elements in **E***Orig* contains the envelope scalefactors of the original signal.

 the output sampling rate from the SBR Tool.

 internal sampling frequency of the SBR Tool, twice the sampling frequency of the core coder (after sampling frequency mapping, ISO/IEC 14496-3:2001, Table 4.55). The sampling frequency of the SBR enhanced output signal is equal to the internal sampling frequency of the SBR Tool, unless the SBR Tool is operated in downsampled mode. If the SBR Tool is operated in downsampled mode, the output sampling frequency is equal to the sampling frequency of the core coder.

 has two column vectors containing the frequency border tables for low and high frequency resolution.

**f***TableHigh* is of length *NHigh*+1 and contains frequency borders for high frequency resolution SBR envelopes.

**f***TableLow* is of length *NLow*+1 and contains frequency borders for low frequency resolution SBR envelopes.

*LE* number of SBR envelopes.

*LQ* number of noise floors.

NQ number of noise floor bands.

 number of frequency bands for low and high frequency resolution.

*numTimeSlots* number of SBR envelope time slots that exist within an AAC frame, 16 for a 1024 AAC frame and 15 for a 960 AAC frame.

 offset-values for the SBR envelope and noise floor data, when using coupled channels.

**Q***Orig* has *LQ* columns where each column is of length NQ and contains the noise floor scalefactors.

 frequency resolution for all SBR envelopes in the current SBR frame, zero for low resolution, one for high resolution.

**t***E* is of length *LE*+1 and contains start and stop time borders for all SBR envelopes in the current SBR frame.

**t***Q* is of length *LQ*+1 and contains start and stop time borders for all noise floors in the current SBR frame.

**Y** is the complex output QMF bank subband matrix from the HF adjuster.

## 3.3 Abbreviations

For the purposes of this TS, the following abbreviations apply.

AAC Advanced Audio Coding

aacPlus Combination of MPEG-4 AAC and MPEG-4 Bandwidth extension (SBR)

Enhanced aacPlus Combination of MPEG-4 AAC, MPEG-4 Bandwidth extension (SBR) and MPEG-4 Parametric Stereo

MPEG Moving Picture Experts Group

SBR Spectral Band Replication

# 4 Outline description

This TS is structured as follows:

Section 5 gives a detailed description of the error concealment algorithms in the Enhanced aacPlus decoder. In Section 5.1 the error concealment of the AAC is described, and in section 5.2 the error concealment of the SBR algorithm is outlined.

Section 6 gives a detailed description of how stereo SBR parameters are down mixed to mono SBR parameters.

Section 7 gives a detailed description of the additional downsampler tool, enabling the Enhanced aacPlus codec to give output sampling rates of 8 and 16kHz, disregarded the sampling rate used for the coded signal.

# 5 Error concealment

## 5.1 AAC error concealment

The AAC core decoder includes a concealment function that increases the delay of the decoder by one frame.

There are various tests inside the core decoder, starting with simple CRC tests and ending in a variety of plausibility checks. If such a check indicates an invalid bitstream, then concealment is applied. Concealment is also applied when the calling main program indicates a distorted or missing data frame using the frameOK flag. This is used for error detection on the transport layer.

Concealment works on the spectral data just before the final frequency to time conversion. In case a single frame is corrupted, concealment interpolates between the last good and the first good frame to create the spectral data for the missing frame. Always the previous frame will be processed by the frequency to time conversion, so here the missing frame to be replaced is the previous frame, the last good frame is the frame before the previous one and the first good frame is the actual frame. If multiple frames are corrupted, concealment implements first a fade out based on slightly modified spectral values from the last good frame. As soon as good frames are available, concealment fades in the new spectral data.

**Interpolation of one corrupt frame:**

In the following the actual frame is frame number *n*, the corrupt frame to be interpolated is the frame *n-*1 and the last but one frame has the number *n-*2.

The determination of window sequence and the window shape of the corrupt frame follows from the table below:

Table 1: Interpolated window sequences and window shapes

|  |  |  |  |
| --- | --- | --- | --- |
| window sequence *n-2* | window sequence *n* | window sequence *n*-1 | window shape *n*-1 |
| ONLY\_LONG\_SEQUENCE or LONG\_START\_SEQUENCE or LONG\_STOP\_SEQUENCE | ONLY\_LONG\_SEQUENCE or LONG\_START\_SEQUENCE or LONG\_STOP\_SEQUENCE | ONLY\_LONG\_SEQUENCE | 0 |
| ONLY\_LONG\_SEQUENCE or LONG\_START\_SEQUENCE or LONG\_STOP\_SEQUENCE | EIGHT\_SHORT\_SEQUENCE | LONG\_START\_SEQUENCE | 1 |
| EIGHT\_SHORT\_SEQUENCE | EIGHT\_SHORT\_SEQUENCE | EIGHT\_SHORT\_SEQUENCE | 1 |
| EIGHT\_SHORT SEQUENCE | ONLY\_LONG\_SEQUENCE or LONG\_START\_SEQUENCE or LONG\_STOP\_SEQUENCE | LONG\_STOP\_SEQUENCE | 0 |

The scalefactor band energies of frames *n*-2 and *n* are calculated. If the window sequence in one of these frames is an EIGHT\_SHORT\_SEQUENCE and the final window sequence for frame *n*-1 is one of the long transform windows, the scalefactor band energies are calculated for long block scalefactor bands by mapping the frequency line index of short block spectral coefficients to a long block representation. The new interpolated spectrum is built by reusing the spectrum of the older frame *n*-2 multiplying a factor to each spectral coefficient. An exception is made in the case of a short window sequence in frame *n*-2 and a long window sequence in frame *n*, here the spectrum of the actual frame n is modified by the interpolation factor. This factor is constant over the range of each scalefactor band and is derived from the scalefactor band energy differences of frames *n*-2 and *n*. Finally the sign of the interpolated spectral coefficients will be flipped randomly.

**Fade out and in:**

A complete fading out takes 5 frames. The spectral coefficients from the last good frame are copied and attenuated by a factor of:



with  as frame counter since the last good frame.

After 5 frames of fading out the concealment switches to muting, that means the complete spectrum will be set to 0.

The decoder fades in when receiving good frames again. The fade in process takes 5 frames, too and the factor multiplied to the spectrum is:



where  is the frame counter since the first good frame after concealing multiple frames.

## 5.2 SBR error concealment

The SBR error concealment algorithm is based on using previous envelope and noise-floor values with an applied decay, as a substitute for the corrupt data. In the flowchart of Figure 1 the basic operation of the SBR error concealment algorithm is outlined. If the frame error flag is set, error concealment bitstream data is generated to be used instead of the corrupt bitstream data. The concealment data is generated according to the following.

The time frequency grids are set to:















The delta coding direction for both the envelope data and noise-floor data are set to be in the time-direction. The envelope data is calculated according to:



where





And where *bs\_amp\_res* and *bs\_coupling* are set to the values of the previous frame.

The noise floor data is calculated according to:



Furthermore, the inverse-filtering levels in *bs\_invf\_mode* are set to the values of the previous frame, and all elements in *bs\_add\_harmonic* are set to zero.

If the frame error is not set, the present time grid and envelope data may need modification if the previous frame was corrupt. If the previous frame was corrupt the time grid of the present frame is modified in order to make sure that there is a continuous transition between the frames. The envelope data for the first envelope is modified according to:



where

.

After the delta coded data has been decoded, a plausibility check is performed to make sure that the decoded data is within reasonable limits. The required limits are:

For the envelope data the logarithmic values shall fulfil:



otherwise the frame will be considered corrupt.

The time grids are also verified according to the following rules (if any of the below is true the frame is considered to be corrupt):

- 

- 

- 

- 

- 

- 

- 

- 

- 

- 

- 

- 

- 

- 

- all elements of **t***Q* are not among the elements of **t***E*

If the plausibility check fails, the frame error flag is set and the error concealment outlined above is applied.



Figure 1: SBR error concealment overview

# 6 SBR stereo parameter to mono parameter downmix

This module enables a decoder only capable of mono output to downmix stereo SBR parameters to mono parameters, hence mono decoding can be performed instead of a stereo decoding.

For all the descriptions in this section left and right applies to the two stereo channels that are being mapped to one channel from here on called merged.

## 6.1 Inverse filtering

The inverse filtering is mapped from stereo to mono by taking the maximum value from left or right as shown below:



## 6.2 Additional harmonics

The additional sinusoids in the merged mono bitstream are the union of the additional sinusoids present in left or right channel as shown below.



## 6.3 Envelope time borders

The merging of the time envelope grid is explained by using  which contains all start/stop borders for all SBR envelopes in the current SBR frame, and the creation of  from the bitstream is explained in [1].

Merging the two  into one is done in the following sequence

1. The start border value of the merged  is set to the largest of the start border values for the left and right channels.

2. The union of all borders for left and right channel values larger than the merged start border is added to the merged 

3. All envelopes smaller than two time slots are removed from the merged . This is achieved by starting from the beginning of  and remove all stop borders that are closer than 2 from the start border.

4. If there are more than 5 envelopes the number of envelopes are reduced. This is achieved by starting from the end of  and search towards the beginning of  for a envelope smaller than 4 and remove the start border of that envelope, this continues until there are 5 envelopes left.

The index  defined in Table 4.119 in [1] has to be merged into one from two. The merging algorithm uses the following recursion:



## 6.4 Noise time borders

The number of noise floors  and the start/stop borders for the noise  is calculated according to [1] for left and right channel and then merged as shown below.



After this is done. It is possible that the middle noise floor border dose not coincide with any SBR envelope time border. If this is the case the middle noise floor border is set equal to the closest time envelope border in the upwards direction.

## 6.5 Envelope data

Before merging the actual data the per envelope frequency selection of high and low frequency resolution is merged as shown below.

,

where  is defined by  and  is defined by 

The envelope data referred to as  in [1] for the left and right channels are merged into  as shown below.

where  is defined by  and

 is defined by  and

 is defined by  and

 is defined by .

## 6.6 Noise floor data

The noise floor data is merged as the average between left and right channel data according to the function below.

,

where  is defined by  and  is defined by .

# 7 Output resampler tool

The audio output from an Enhanced aacPlus decoder using a downsampled synthesis filterbank (see 4.6.18.4.3 or 4.6.18.8.2.3 of [1]) is given at the AAC decoder sampling rate . If an output with lower sampling rate is desired, a downsampler tool is required in order to convert the audio output from the source sampling frequency  to the target sampling frequency .

The downsampler tool consists of three parts connected to the final operations of the SBR decoder as in Figure 1. (Refering to Figure 4.44 of [1]). The QMF bandlimiter operates on QMF samples and is inserted prior to the QMF synthesis bank, the spline resampler and the postfilter operate on synthesised PCM audio samples.



Figure 1: Downsampler tool

## 7.1 QMF bandlimiter

The QMF bandlimiter maps the final QMF matrix to , according to the rule



where the cutoff frequency subband indexis given by



## 7.2 Spline resampler

Given the discrete time output  from synthesis QMF bank the spline resampler relies on the continuous time signal representation,



where is a cubic B-spline supported on the interval .The discrete time output of the resampler is then defined, up to a suitable delay, by . For each , define the integer and the remainder by



Then the spline evaluation results in the time varying causal filtering,



with homogeneous initialization and filter coefficients



## 7.3 Postfilter

In order to compensate for a small loss in signal power for high frequencies, a first order IIR postfilter is applied to the output of the spline resampler, resulting in the final output , where  and



The value of the parameter is given by Table 1.

Table 1: Postfilter coefficient 

|  |  |  |
| --- | --- | --- |
|  |  | |
| 8 kHz | 16 kHz |
| 22.05 kHz | 0.06 | 0.30 |
| 24 kHz | 0.05 | 0.24 |

Annex A (informative):  
Change history

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Change history** | | | | | | | |
| **Date** | **TSG SA#** | **TSG Doc.** | **CR** | **Rev** | **Subject/Comment** | **Old** | **New** |
| 2004-09 | 25 | SP-040634 |  |  | Approved at SA#25 | 2.0.0 | 6.0.0 |
| 2005-09 | 29 | SP-050426 | 0001 |  | Correction of written specification: AAC concealment fade out factor | 6.0.0 | 6.1.0 |
| 2007-06 | 36 |  |  |  | Version for Release 7 | 6.1.0 | 7.0.0 |
| 2008-12 | 42 |  |  |  | Version for Release 8 | 7.0.0 | 8.0.0 |
| 2009-12 | 46 |  |  |  | Version for Release 9 | 8.0.0 | 9.0.0 |
| 2011-03 | 51 |  |  |  | Version for Release 10 | 9.0.0 | 10.0.0 |
| 2012-09 | 57 |  |  |  | Version for Release 11 | 10.0.0 | 11.0.0 |
| 2014-09 | 65 |  |  |  | Version for Release 12 | 11.0.0 | 12.0.0 |
| 2015-12 | 70 |  |  |  | Version for Release 13 | 12.0.0 | 13.0.0 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Change history** | | | | | | | |
| **Date** | **Meeting** | **TDoc** | **CR** | **Rev** | **Cat** | **Subject/Comment** | **New version** |
| 2017-03 | 75 |  |  |  |  | Version for Release 14 | 14.0.0 |
| 2018-06 | 80 |  |  |  |  | Version for Release 15 | 15.0.0 |
| 2020-07 | - | - | - | - | - | Update to Rel-16 version (MCC) | **16.0.0** |