INTERNATIONAL STANDARD

# ISO/IEC 23003-3

First edition
2012-04-01

**AMENDMENT 1**
2014-03-15

# Information technology — MPEG audio technologies —

## Part 3: Unified speech and audio coding

## AMENDMENT 1: Conformance



*Technologies de l'information — Technologies audio MPEG —*

*Partie 3: Discours unifié et codage audio*

*AMENDEMENT 1: Conformité*



Reference number
ISO/IEC 23003-3:2012/Amd.1:2014(E)

# Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Amendment 1 to ISO/IEC 23003-3:2012 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

# Information technology — MPEG audio technologies —

## Part 3:
## Unified speech and audio coding

## AMENDMENT 1: Conformance

*In Clause 2, "Normative References", add the following entry:*

ISO/IEC 14496-26:2010, *Information technology — Coding of audio-visual objects — Part 26: Audio conformance*

*In 4.5.4 replace:*

Four different hierarchical levels are defined with increasing number of audio channels and increasing complexity. All four levels include Level 2 of the Baseline USAC profile. The definition of the four levels of the Extended HE AAC profile is given in Table 3. All notes in Table 3 and all restrictions listed in the columns 2, 3, 4, and 5 ("Max. channels/object", "Max. AAC sampling rate, SBR not present [kHz]", "Max. AAC sampling rate, SBR present [kHz]", "Max. SBR sampling rate [kHz] (in/out)") of Table 3 apply only when decoding HE AAC v2 profile compliant bit streams.



**Table 3 — Levels for the Extended HE AAC profile**



| Level (NOTE 1) | Max. channels / object | Max. AAC sampling rate, SBR not present [kHz] | Max. AAC sampling rate, SBR present [kHz] | Max. SBR sampling rate [kHz] (in/out) | Max. PCU | Max. RCU | Max. PCU HQ / LP SBR (NOTE 5) | Max. RCU HQ / LP SBR (NOTE 5) |
|---|---|---|---|---|---|---|---|---|
| 1 | NA | NA | NA | NA | NA | NA | NA | NA |
| 2 | 2 | 48 | 24 | 24/48 | 12 | 11 | 12 | 11 |
| 3 | 2 | 48 | 24/48 (NOTE 3) | 48/48 (NOTE 2) | 15 | 11 | 15 | 11 |
| 4 | 5 | 48 | 24/48 (NOTE 4) | 48/48 (NOTE 2) | 25 | 28 | 20 | 23 |
| 5 | 5 | 96 | 48 | 48/96 | 49 | 28 | 39 | 23 |

NOTE 1: Level 2, 3, and 4 Extended HE AAC profile decoders implement the baseline version of the parametric stereo tool. A level 5 decoder shall not be limited to the baseline version of the parametric stereo tool.

NOTE 2: For level 3 and level 4 decoders, it is mandatory to operate the SBR tool in downsampled mode if the sampling rate of the AAC core is higher than 24kHz. Hence, if the SBR tool operates on a 48kHz signal, the internal sampling rate of the SBR tool will be 96kHz, however, the output signal will be downsampled by the SBR tool to 48kHz.

NOTE 3: If Parametric Stereo data are present the maximum AAC sampling rate is 24kHz, if Parametric Stereo data are not present the maximum AAC sampling rate is 48kHz.

NOTE 4: For one or two channels the maximum AAC sampling rate, with SBR present, is 48kHz. For more than two channels the maximum AAC sampling rate, with SBR present, is 24kHz.

NOTE 5: The PCU/RCU number are given for a decoder operating the LP SBR tool whenever applicable.

*with:*

A number of hierarchical levels are defined with increasing number of audio channels and increasing complexity. All levels include Level 2 of the Baseline USAC profile. The definition of the levels of the Extended HE AAC profile is given in Table 3. All notes in Table 3 and all restrictions listed in the columns 2, 3, 4, and 5 ("Max. channels/object", "Max. AAC sampling rate, SBR not present [kHz]", "Max. AAC sampling rate, SBR present [kHz]", "Max. SBR sampling rate [kHz] (in/out)") of Table 3 apply only when decoding HE AAC v2 profile compliant bit streams.

**Table 3 — Levels for the Extended HE AAC profile**

| Level (NOTE 1) | Max. channels / object | Max. AAC sampling rate, SBR not present [kHz] | Max. AAC sampling rate, SBR present [kHz] | Max. SBR sampling rate [kHz] (in/out) | Max. PCU | Max. RCU | Max. PCU HQ / LP SBR (NOTE 5) | Max. RCU HQ / LP SBR (NOTE 5) |
|---|---|---|---|---|---|---|---|---|
| 1 | NA | NA | NA | NA | NA | NA | NA | NA |
| 2 | 2 | 48 | 24 | 24/48 | 12 | 11 | 12 | 11 |
| 3 | 2 | 48 | 24/48 (NOTE 3) | 48/48 (NOTE 2) | 15 | 11 | 15 | 11 |
| 4 | 5 | 48 | 24/48 (NOTE 4) | 48/48 (NOTE 2) | 25 | 28 | 20 | 23 |
| 5 | 5 | 96 | 48 | 48/96 | 49 | 28 | 39 | 23 |
| 6 | 7 | 48 | 24/48 (NOTE 4) | 48/48 | 34 | 37 | 27 | 30 |
| 7 | 7 | 96 | 48 | 48/96 | 67 | 37 | 53 | 30 |

NOTE 1: Level 2, 3, 4, 6 and 7 Extended HE AAC profile decoders implement the baseline version of the parametric stereo tool. A level 5 decoder shall not be limited to the baseline version of the parametric stereo tool.

NOTE 2: For level 3, 4 and 6 decoders, it is mandatory to operate the SBR tool in downsampled mode if the sampling rate of the AAC core is higher than 24kHz. Hence, if the SBR tool operates on a 48kHz signal, the internal sampling rate of the SBR tool will be 96kHz, however, the output signal will be downsampled by the SBR tool to 48kHz.

NOTE 3: If Parametric Stereo data are present the maximum AAC sampling rate is 24kHz, if Parametric Stereo data are not present the maximum AAC sampling rate is 48kHz.

NOTE 4: For one or two channels the maximum AAC sampling rate, with SBR present, is 48kHz. For more than two channels the maximum AAC sampling rate, with SBR present, is 24kHz.

NOTE 5: The PCU/RCU number are given for a decoder operating the LP SBR tool whenever applicable.

NOTE 6: A Level 6 or 7 decoder is not required to decode a Level 5 stream.



*In 5.3.2 amend Table 36 as follows:*

 

**Table 36 — Syntax of acelp_coding()**

| Syntax | No. of bits | Mnemonic |
|---|---|---|
| acelp_coding(acelp_core_mode) | | |
| { | | |
| [...] | | |
| switch (acelp_core_mode) { | | |
| case 0 | | |
| **icb_index**[sfr]; | **20** | **uimsbf** |
| break; | | |
| case 1 | | |
| **icb_index**[sfr]; | **28** | **uimsbf** |
| break; | | |
| case 2 | | |
| **icb_index**[sfr]; | **36** | **uimsbf** |
| break; | | |
| case 3 | | |
| **icb_index**[sfr]; | **44** | **uimsbf** |
| break; | | |
| case 4 | | |
| **icb_index**[sfr]; | **52** | **uimsbf** |
| break; | | |
| case 5 | | |
| **icb_index**[sfr]; | **64** | **uimsbf** |
| break; | | |
| case 6 | | |
| **icb_index**[sfr]; | **12** | **uimsbf** |
| break; | | |
| case 7 | | |
| **icb_index**[sfr]; | **16** | **uimsbf** |
| break; | | |
| } | | |
| **gains[sfr];** | **7** | **uimsbf** |
| } | | |
| NOTE: coreCoderFrameLength designates the core frame length in samples and is equal to either 1024 or 768. See also 6.1.1.2. | | |



*In 7.14.5.2.1, replace:*

Depending on the coding mode, the following codebooks are used:

*with:*

Depending on the coding mode, the following codebooks are used:

— 12-bit codebook with 2 pulses $i_0$ and $i_1$. Pulse $i_0$ can be selected from either track 0 or 2, pulse $i_1$ can be selected from either track 1 or 3 (5×2+2)

 

— 16-bit codebook with 3 pulses on three tracks. One pulse on track 0, one pulse on track 2 and one pulse on either track 1 or 3 (selected track signalled by a 1 bit field), which amounts to (5×3+1) = 16 bits.

*Add a new Clause 8, "Conformance testing", as shown below:*

# 8 Conformance testing

## 8.1 Introduction

The present Clause 8 specifies conformance criteria for both bitstreams and decoders compliant with the USAC standard as defined in this document. This is done to assist implementers and to ensure interoperability.

## 8.2 Terms and definitions

**bitstream**
encoded audio data

**conformance data**
conformance test sequences and conformance tools

**conformance tool**
tool to check certain conformance criteria



**conformance test sequence**
generic term for conformance test bitstreams and corresponding reference waveforms

**conformance test bitstream**
USAC bitstream used for testing the conformance of a USAC decoder



**conformance test condition**
condition which applies to properties of a conformance test bitstream in order to test a certain functionality of the USAC decoder

**conformance test case**
combination of one or more conformance test conditions for which a set of conformance test sequences is provided

**main audio channel**
audio channel conveyed by means of a UsacSingleChannelElement or UsacChannelPairElement

**reference waveform**
decoded counterpart of a bitstream

**USAC bitstream**
data encoded according to the USAC standard

**UsacCPE**
UsacChannelPairElement

**UsacEXT**
UsacExtElement

**UsacLFE**
UsacLfeElement

**UsacSCE**
UsacSingleChannelElement

## 8.3 USAC conformance testing

### 8.3.1 Profiles

Profiles are defined in 4.5. Some conformance criteria apply to USAC in general, while others are specific to certain profiles and their respective levels. Conformance shall be tested for the level of the profile with which a given bitstream or decoder claims to comply.

In addition to the conformance requirements described in this clause, a decoder which claims to comply with the Extended HE AAC Profile shall fulfill conformance for the HE AAC v2 profile according to ISO/IEC 14496-26:2010.

### 8.3.2 Conformance tools and test procedure

To test USAC compliant audio decoders, ISO/IEC JTC 1/SC 29/WG 11 supplies a number of conformance test sequences. Supplied sequences cover all profiles as defined in 4.5. For a supplied test sequence, testing can be done by comparing the output of a decoder under test with a reference waveform also supplied by ISO/IEC JTC 1/SC 29/WG 11. In cases where the decoder under test is followed by additional operations (e.g. quantizing a signal to a 16 bit output signal) the conformance point is prior to such additional operations, i.e. it is permitted to use the actual decoder output (e.g. with more than 16 bit) for conformance testing.

Measurements are carried out relative to full scale where the output signals of the decoders are normalized to be in the range between −1.0 and +1.0.



In ISO/IEC 14496-26:2010 a set of test methods is defined to test the output of the decoder under test against the reference output. RMS/LSB Measurement, Segmental SNR and PNS conformance criteria are used for the comparison. A particular test method for a certain test sequence is specified in 8.5.



For elements producing output that cannot be tested with the methods described in ISO/IEC 14496-26:2010, specific conformance testing procedures are described in 8.5.



#### 8.3.2.1 Conformance data

All test sequences are provided in the shape of a zip archive as an electronic attachment. Furthermore, an MS Excel worksheet (“Usac_Conformance_Tables.xlsx”) is provided as an electronic attachment that lists all test sequences for each module.

For all conformance test sequences, the file names are composed of several parts which convey information about:

— which module of the decoder is tested

— which channelConfigurationIndex is employed

— which test conditions apply to the test sequence

— which coreSbrFrameLengthIndex applies to the test sequence

— which sampling frequency is signalled in the test sequence

The file naming convention given in Table 149 is used. Values in box brackets are optional.

**Table 149 — File name conventions**

| Module | File Name (compressed) | File Name (uncompressed) |
|---|---|---|
| Frequency domain coding (FD mode), 8.4.4 | Fd_<cCI>_c<cSFLI>_<testCase>_<uSFI>.mp4 | FD_<cCI>_c<cSFLI>_<testCase>_<uSFI>.wav |
| Linear predictive domain coding (LPD mode), 8.4.5 | Lpd_<cCI>_c<cSFLI>_<testCase>_[<uSFI>].mp4 | Lpd_<cCI>_c<cSFLI>_<testCase>_[<uSFI>].wav |
| Common core coding tools, 8.4.6 | Cct_<cCI>_c<cSFLI>_<testCase>_[<uSFI>].mp4 | Cct_<cCI>_c<cSFLI>_<testCase>_[<uSFI>].wav |
| Enhanced spectral band replication (eSBR), 8.4.7 | eSbr_<cSFLI>_<testCase>.mp4 | eSbr_<cSFLI>_<testCase>.wav |
| MPEG Surround 2-1-2, 8.4.10 | Mps_<bsFR>_Sc<sCI>_<testCase>.mp4 | Mps_<bsFR>_Sc<sCI>_<testCase>.wav |

**<cCI>** channelConfigurationIndex as described in Table 68.

**<testCase>** Setup string. May consist of a concatenation of one or more abbreviations as listed in Table 150. If no setup string is specified the basic test conditions apply

**<cSFLI>** coreSbrFrameLengthIndex as described in Table 70.



**<uSFI>** usacSamplingFrequencyIndex as described in Table 67. If the escape value is specified the used sampling frequency is appended, e.g. “xx_1f_42000.mp4” for a sampling frequency of 42 kHz.



**<bsFR>** bsFreqRes as described in ISO/IEC 23003-1:2007, Table 39

**<sCI>** stereoConfigIndex as described in Table 72

**Table 150 — Test conditions and abbreviations**

| FD core mode | |
|---|---|
| **Test Condition** | **Abbrev.** |
| FD window switching test condition | Win |
| Noise filling test condition | Nf |
| Tns test condition | Tns |
| Varying max_sfb test condition | Sfb |
| Handling of extensions condition | Ex |
| Arithmetic coder test condition | Ac |
| Non-meaningful FD window switching test condtion | Nmf |
| M/S stereo test condition | Ms |
| Complex prediction stereo test condition | Cp |

| LPD core mode | |
|---|---|
| **Test Condition** | **Abbrev.** |
| LPC coding test condition | Lpc |
| ACELP core mode test condition | Ace |
| TCX and noise filling test condition | Tcx |
| LPD mode coverage and FAC test condition | Lpd |
| Bass-post filter test condition | Bpf |
| AVQ test condition | Avq |

| Combined core coding | |
|---|---|
| **Test Condition** | **Abbrev.** |
| FD-LPD transition and FAC test condition | Flt |
| FD/TCX noise filling test condition | Cnf |
| Bass-post filter test condition | Cbf |
| synchr. FD-LPD transition and FAC test condition | Flts |
| asynchr. FD-LPD transition and FAC test condition | Flta |
| Arithmetic coder test condition | CAc |



| eSbr | |
|---|---|
| **Test Condition** | **Abbrev.** |
| QMF accuracy test condition | Qma |
| Envelope adjuster accuracy and SBR pre-processing test condition | Eaa |
| Header and grid control test condition test condition | Hgt |
| Inverse filtering test condition | Ift |
| Additional sine test (missing harmonics) test condition | Ast |
| Sampling rate test condition | Sr |
| Channel mode test condition | Cm |
| interTes test condition | Tes |
| PVC test condition | Pvc |
| Harmonic transposition (QMF) test condition | Htq |
| Harmonic transposition (crossproducts) test condition | Xp |
| Transposer toggle test condition | Ttt |
| Envelope shaping toggle (PVC on/off) test condition | Est |
| Varying crossover frequency test condition | Xo |
| stereoConfigIndex test condition | Mps |

 

| Mpeg surround 212 | |
|---|---|
| **Test Condition** | **Abbrev.** |
| TSD test condition | Tsd |
| Rate mode test condition | Rm |
| Phase coding test condition | Pc |
| Decorrelator configuration. test condition | Dc |
| DMX gain test condition | Dm |
| Bands phase test condition | Bp<X> |
| Pseudo lr test condition | Plr |
| Residual bands test condition | Rb<X> |

## 8.4 USAC Bitstreams

### 8.4.1 General

#### 8.4.1.1 Characteristics

Characteristics of bitstreams specify the constraints that are applied by the encoder in generating the bitstreams. These syntactic and semantic constraints may, for example, restrict the range or the values of parameters that are encoded directly or indirectly in the bitstreams. The constraints applied to a given bitstreams may or may not be known a priori.



#### 8.4.1.2 Test procedure



Each USAC bitstream shall meet the syntactic and semantic requirements specified in this document. The present subclause defines the conformance criteria that shall be fulfilled by a compliant bitstream. These criteria are specified for the syntactic elements of the bitstream and for some parameters decoded from the USAC bitstream payload.

For each tool a set of semantic tests to be performed on the bitstreams is described. To verify whether the syntax is correct is straightforward and therefore not defined herein after. In the description of the semantic tests it is assumed that the tested bitstreams contains no errors due to transmission or other causes. For each test the condition or conditions that must be satisfied are given, as well as the prerequisites or conditions in which the test can be applied.

### 8.4.2 USAC Configuration

#### 8.4.2.1 Characteristics

Encoders may apply restrictions to the following parameters of the bitstream:

a) usacSamplingFrequencyIndex

b) usacSamplingFrequency

c) coreSbrFrameLengthIndex

d) channelConfigurationIndex

e) presence of configuration extensions

f) numOutChannels

g) bsOutputChannelPos

 

h) numElements

i) stereoConfigIndex

j) use of time warped MDCT

k) use of noise filling in FD mode

l) use of the eSBR harmonic transposer

m) use of the eSBR inter-TES tool

n) use of the eSBR PVC tool

o) SBR default header, for details see 8.4.7.

p) MPS config, for details see 8.4.10.

#### 8.4.2.2 Test procedure

##### 8.4.2.2.1 UsacConfig()

**usacSamplingFrequencyIndex** Shall be encoded with a non-reserved value specified in Table 67. For further profile and level dependent restrictions see 8.4.11.

**usacSamplingFrequency** No restrictions apply. For profile and level dependent restrictions see 8.4.11.



**coreSbrFrameLengthIndex** no restrictions apply

**channelConfigurationIndex** Shall be encoded with a non-reserved value specified in Table 68. For further profile and level dependent restrictions see 8.4.11. In the case of channelConfigurationIndex==0 further restrictions apply as described in 8.4.2.2.2.



**usacConfigExtensionPresent** no restrictions apply

##### 8.4.2.2.2 UsacChannelConfig()

numOutChannels no restrictions apply. For profile and level dependent restrictions see 8.4.11.

**bsOutputChannelPos** A bsOutputChannelPos of value 3 or 26 (LFE speaker positions) shall be associated with an LFE channel. Any other value shall be associated with a main audio channel.

##### 8.4.2.2.3 UsacDecoderConfig()

numElements the value of this data element shall be such that the accumulated sum of all channels contained in the bitstream complies with the restrictions outlined in 8.4.2.2.1.

**usacElementType** no restrictions apply. For profile and level dependent restrictions see 8.4.11.

 

#### 8.4.2.2.4 UsacSingleChannelElementConfig()

No restrictions are applicable to this bitstream element.

#### 8.4.2.2.5 UsacChannelPairElementConfig()

NOTE The UsacChannelPairElementConfig() element and all included elements may only be present when coding more than one output channel (see restrictions applying to UsacConfig() in 8.4.2.2.1).

| | |
|---|---|
| **stereoConfigIndex** | no restrictions apply |

#### 8.4.2.2.6 UsacLfeElementConfig()

No restrictions are applicable to this bitstream element.

#### 8.4.2.2.7 UsacCoreConfig()

| | |
|---|---|
| **tw_mdct** | no restrictions apply. For profile and level dependent restrictions see 8.4.11. |
| **noiseFilling** | no restrictions apply |

#### 8.4.2.2.8 SbrConfig()



| | |
|---|---|
| **harmonicSBR** | no restrictions apply |
| **bs_interTes** | no restrictions apply |
| **bs_pvc** | no restrictions apply |

#### 8.4.2.2.9 SbrDfltHeader()

| | |
|---|---|
| **dflt_start_freq** | no restrictions apply |
| **dflt_stop_freq** | no restrictions apply |
| **dflt_header_extra1** | no restrictions apply |
| **dflt_header_extra2** | no restrictions apply |
| **dflt_freq_scale** | no restrictions apply |
| **dflt_alter_scale** | no restrictions apply |
| **dftl_nose_bands** | no restrictions apply |
| **dflt_limiter_bands** | no restrictions apply |
| **dflt_limiter_gains** | no restrictions apply |
| **dflt_interpol_freq** | no restrictions apply |
| **dflt_smoothing_mode** | no restrictions apply |

#### 8.4.2.2.10 Mps212Config()

| | |
|---|---|
| **bsFreqRes** | shall not be encoded with a value of 0 |
| **bsFixedGainDMX** | no restrictions apply |
| **bsTempShapeConfig** | no restrictions apply |
| **bsDecorrConfig** | shall not be encoded with a value of 3 |
| **bsHighRateMode** | no restrictions apply |
| **bsPhaseCoding** | no restrictions apply |
| **bsOttBandsPhasePresent** | no restrictions apply |
| **bsOttBandsPhase** | shall not be encoded with a value larger than the value of numBands as given by ISO/IEC 23003-1:2007, 5.2, Table 39 and depends on bsFreqRes. |
| **bsResidualBands** | shall not be encoded with a value larger than the value of numBands as given by ISO/IEC 23003-1:2007, 5.2, Table 39 and depends on bsFreqRes. |
| **bsPseudoLr** | no restrictions apply |
| **bsEnvQuantMode** | shall be 0 |



#### 8.4.2.2.11 UsacExtElementConfig()



| | |
|---|---|
| usacExtElementType | no restrictions apply |
| usacExtElementConfigLength | no restrictions apply |
| **usacExtElementDefaultLengthPresent** | no restrictions apply |
| usacExtElementDefaultLength | no restrictions apply |
| **usacExtElementPayloadFrag** | no restrictions apply |

#### 8.4.2.2.12 UsacConfigExtension()

| | |
|---|---|
| numConfigExtensions | no restrictions apply |
| usacConfigExtType[] | no restrictions apply |
| usacConfigExtLength[] | no restrictions apply |
| **fill_byte** | should be '10100101' |

### 8.4.3 Framework

#### 8.4.3.1 Characteristics

Encoders may apply restrictions to the following parameters of the bitstream:

a) signalling of independently decodable frames

 