**3GPP TSG-SA WG4 Meeting #127S4-240344**

**Sophia-Antipolis, France, 29 January - 2 February 2024**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *CR-Form-v12.2* | | | | | | | | |
| **CHANGE REQUEST** | | | | | | | | |
|  | | | | | | | | |
|  | **26.244** | **CR** |  | **rev** |  | **Current version:** | **17.0.0** |  |
|  | | | | | | | | |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* | | | | | | | | |
|  | | | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME |  | Radio Access Network |  | Core Network |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| ***Title:*** | Introduction of IVAS into the 3GPP File Format (3GP) | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Source to WG:*** | Fraunhofer IIS | | | | | | | | | |
| ***Source to TSG:*** | S4 | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Work item code:*** | IVAS\_codec | | | | |  | ***Date:*** | | | 2024-02-01 |
|  |  | | | |  | |  | | |  |
| ***Category:*** | B |  | | | | | ***Release:*** | | | 18 |
|  | *Use one of the following categories:* ***F*** *(correction)* ***A*** *(mirror corresponding to a change in an earlier release)* ***B*** *(addition of feature),* ***C*** *(functional modification of feature)* ***D*** *(editorial modification)*  Detailed explanations of the above categories can be found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | | | | | | | | *Use one of the following releases: Rel-8 (Release 8) Rel-9 (Release 9) Rel-10 (Release 10) Rel-11 (Release 11) … Rel-16 (Release 16) Rel-17 (Release 17) Rel-18 (Release 18) Rel-19 (Release 19)* | |
|  |  | | | | | | | | | |
| ***Reason for change:*** | | The 3GP file format does not support IVAS. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Summary of change:*** | | IVAS and its parameters are specified for the 3GP file format. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Consequences if not approved:*** | | EVS cannot be used in packet-switched streaming services. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Clauses affected:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
|  | | **Y** | **N** |  | | | |  | | |
| ***Other specs*** | |  | **X** | Other core specifications | | | | TS/TR ... CR ... | | |
| ***affected:*** | |  | **X** | Test specifications | | | | TS/TR ... CR ... | | |
| ***(show related CRs)*** | |  | **X** | O&M Specifications | | | | TS/TR ... CR ... | | |
|  | |  | | | | | | | | |
| ***Other comments:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***This CR's revision history:*** | |  | | | | | | | | |

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TS 22.233: "Transparent End-to-End Packet-switched Streaming Service; Stage 1".

[2] 3GPP TS 26.233: "Transparent end-to-end packet switched streaming service (PSS); General description".

[3] 3GPP TS 26.234: "Transparent end-to-end packet switched streaming service (PSS); Protocols and codecs".

[4] 3GPP TS 26.245: "Transparent end-to-end packet switched streaming service (PSS); Timed text format".

[5] 3GPP TS 26.246: "Transparent end-to-end packet switched streaming service (PSS); 3GPP SMIL Language Profile".

[6] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[7] ISO/IEC 14496-12 | 15444-12: "Information technology - Coding of audio-visual objects - Part 12: ISO base media file format" | "Information technology - JPEG 2000 image coding system - Part 12: ISO base media file format".

[8] 3GPP TS 26.140: "Multimedia Messaging Service (MMS); Media formats and codecs".

[9] ITU-T Recommendation H.263 (01/05): "Video coding for low bit rate communication".

[10] ISO/IEC 14496-2:2004: "Information technology – Coding of audio-visual objects – Part 2: Visual".

[11] 3GPP TS 26.071: "Mandatory Speech CODEC speech processing functions; AMR Speech CODEC; General description".

[12] 3GPP TS 26.171: "AMR Wideband Speech Codec; General Description".

[13] ISO/IEC 14496-3:2005: "Information technology – Coding of audio-visual objects – Part 3: Audio".

[14] ISO/IEC 14496-14:2003: "Information technology – Coding of audio-visual objects – Part 14: MP4 file format".

[15] IETF RFC 4867: " RTP Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", Sjoberg J. et al., April 2007.

[16] 3GPP TS 26.101: "Mandatory Speech Codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec frame structure".

[17] 3GPP TS 26.201: "Speech Codec speech processing functions; AMR Wideband Speech Codec; Frame Structure".

[18] void

[19] IETF RFC 3711: "The Secure Real-time Transport Protocol", Baugher M. et al., March 2004.

[20] ISO/IEC 14496-15: "Information technology – Coding of audio-visual objects – Part 15: Carriage of NAL unit structured video in the ISO base media file format".

[21] 3GPP TS 26.290: "Extended AMR Wideband codec; Transcoding functions".

[22] void

[23] 3GPP TS 26.401: "General audio codec audio processing functions; Enhanced aacPlus general audio codec; General description".

[24] 3GPP TS 26.410: "General audio codec audio processing functions; Enhanced aacPlus general audio codec; Floating-point ANSI-C code".

[25] 3GPP TS 26.411: "General audio codec audio processing functions; Enhanced aacPlus general audio codec; Fixed-point ANSI-C code".

[26] void

[27] IETF RFC 3839: "MIME Type Registrations for 3rd Generation Partnership Project (3GPP) Multimedia files", Castagno R. and Singer D., July 2004.

[28] IETF RFC 4396: "RTP Payload Format for 3rd Generation Partnership Project (3GPP) Timed Text", Rey J. and Matsui Y., February 2006.

[29] ITU-T Recommendation H.264 (04/2013): "Advanced video coding for generic audiovisual services".

[30] IETF RFC 6184: "RTP Payload Format for H.264 Video", Wang Y.-K. et al, May 2011.

[31] IETF RFC 4234: "Augmented BNF for Syntax Specifications: ABNF", Crocker D. and Overell P., October 2005.

[32] MP4REG, MP4 Registration Authority, [www.mp4ra.org](http://www.mp4ra.org).

[33] ID3v2, <http://www.id3.org/>.

[34] IETF RFC 6381: "The 'Codecs' and 'Profiles' Parameters for ``Bucket´´ Media Types", Gellens R., Singer D. and Frojdh P., August 2011.

[35] IETF RFC 4648: "The Base16, Base32, and Base64 Data Encodings", Josefsson S., October 2006.

[36] 3GPP TS 26.142: "Dynamic and Interactive Multimedia Scene".

[37] Open Mobile Alliance: "DRM Specification 2.0".

[38] ISO/IEC 14496-12:2008/PDAM1: "Part 12: ISO base media file format/AMENDMENT 1: General improvements including hint tracks, metadata support, and sample groups".

[39] 3GPP TS 33.246: "Security of Multimedia Broadcast/Multicast Service (MBMS)".

[40] 3GPP TS 26.346: "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs".

[41] void

[42] IETF RFC 4288: "Media Type Specifications and Registration Procedures", Freed N. and Klensin J., December 2005.

[43] IETF RFC 5322: "Internet Message Format", Resnick, P. October 2008.

[44] IETF RFC 5234: "Augmented BNF for Syntax Specifications: ABNF", Crocker D., Overell, P., January 2008.

[45] IETF RFC 2045: "Multipurpose Internet Mail Extensions, (MIME) Part One: Format of Internet Message Bodies", Freed, N., Borenstein, N., November 1996.

[46] IETF RFC 3926: "FLUTE - File Delivery over Unidirectional Transport", Paila T., Luby M., Lehtonen R., Roca V., and Walsh R., October 2004.

[47] void

[48] IETF RFC 2616: "Hypertext Transfer Protocol – HTTP/1.1", Fielding R. et al., June 1999.

[49] 3GPP TS 26.247: "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)".

[50] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".

[51] Recommendation ITU-T H.265 (12/2016): "High efficiency video coding" | ISO/IEC 23008-2:2015: "High Efficiency Coding and Media Delivery in Heterogeneous Environments – Part 2: High Efficiency Video Coding".

[52] IETF RFC 7798 (2016): "RTP Payload Format for High Efficiency Video Coding (HEVC)", Y.-K. Wang, Y. Sanchez, T. Schierl, S. Wenger, M. M. Hannuksela.

[53] ISO/IEC 23001-10:2015 "Information technology – MPEG systems technologies -- Part 10: Carriage of Timed Metadata Metrics of Media in ISO Base Media File Format".

[54] 3GPP TR 26.938: "Packet-switched Streaming Service (PSS); Improved support for dynamic adaptive streaming over HTTP in 3GPP".

[55] 3GPP TS 26.445: "Codec for Enhanced Voice Services (EVS); Detailed Algorithmic Description".

[56] 3GPP TS 26.116: "TV Video Profiles".

[57] 3GPP TS 26.253: "Codec for Immersive Voice and Audio Services (IVAS); Detailed Algorithmic Description and RTP Payload Format".

[58] ISO/IEC 23091-3:2018 "Information technology – Coding-independent code points Part 3: Audio".

[59] 3GPP TS 26.258: " Codec for Immersive Voice and Audio Services (IVAS); C code (floating-point)".

## 3.2 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [6] and the following apply.

3GP 3GPP file format

AAC Advanced Audio Coding

AMR Adaptive Multi-Rate Codec

AMR-WB Adaptive Multi-Rate Wideband Codec

AMR-WB+ Extended Adaptive Multi-Rate Wideband Codec

AVC Advanced Video Coding

ADU Application Data Unit

BIFS Binary Format for Scenes

CVO Coordination of Video Orientation

DIMS Dynamic and Interactive Multimedia Scenes

Enhanced aacPlus MPEG-4 High Efficiency AAC plus MPEG-4 Parametric Stereo

EVS Enhanced Voice Services

FLUTE File Delivery over Unidirectional Transport

HEVC High Efficiency Video Coding

HTTP HyperText Transport Protocol

ITU-T International Telecommunications Union – Telecommunications

IVAS Immersive Voice and Audio Services

MIKEY Multimedia Internet KEYing

MIME Multipurpose Internet Mail Extensions

MMS Multimedia Messaging Service

MP4 MPEG-4 file format

MPD Media Presentation Description

PSS Packet-switched Streaming Service

RAP Random Access Point

RTP Real-time Transport Protocol

RTSP Real-Time Streaming Protocol

SDP Session Description Protocol

SRTP Secure Real-time Transport Protocol

URL Uniform Resource Locator

### 5.2.2 Registration of codecs

Code streams for H.263 video [9], H.264 (AVC) video [29], H.265 (HEVC) video [51], AMR narrow-band speech [11], AMR wide-band speech [12], Extended AMR wide-band audio [21], EVS [55], IVAS [57], Enhanced aacPlus audio [23, 24, 25], MPEG-4 AAC audio [13], and timed text [4] can be included in 3GP files as described in clause 6 of the present document.

## 6.1 General

The purpose of this clause is to define the necessary structure for integration of the H.263, AMR, AMR-WB, Extended AMR-WB (AMR-WB+), EVS, IVAS, Enhanced aacPlus and AAC media specific information in a 3GP file. Clause 6.2 gives some background information about the Sample Description box in the ISO base media file format [7] and clause 6.4 about the MP4AudioSampleEntry box in the MPEG-4 file format [14]. The definitions of the Sample Entry boxes for AMR, AMR-WB, AMR-WB+ and H.263 are given in clauses 6.5 to 6.10. The definition of the Sample Entry box for EVS is given in clause 6.14. The integration of timed text in a 3GP file is specified in [4], the integration of H.264 (AVC) is specified in [20], the integration of H.265 (HEVC) is specified in clause 8 of [20], the integration of Quality metrics timed metadata track is specified in clause 4 of [53] and clause 16 of this specification and the integration of DIMS is specified in [36] and clauses 5.4.3, 5.4.6 and 11 of the present document. Requirements for integrating video codecs in the context of the TV Video Profile are documented in TS 26.116 [56].

AMR and AMR-WB data is stored in the stream according to the AMR and AMR-WB storage format for single channel header of Annex E [15], without the AMR magic numbers. The 3GPP file format is the native storage format for AMR-WB+. The data stream, stored in samples of a 3GP file, shall be formatted according to clause 8.3 of [21]. Each sample contains one or more AMR-WB+ storage units. The number of storage units per sample may differ from sample to sample.

For EVS each sample of the media is one speech frame block as specified in Annex A.2.6.2 of [55]. A speech frame block consists of N ToC entries and N speech frames, where N is the value of channelcount in the EVSSampleEntry box specified in clause 6.14 of the present document.

## 6.2 Sample Description box

In an ISO file, Sample Description Box gives detailed information about the coding type used, and any initialisation information needed for that coding. The Sample Description Box can be found in the ISO file format Box Structure Hierarchy shown in figure 6.1.



Figure 6.1: ISO File Format Box Structure Hierarchy

The Sample Description Box can have one or more Sample Entries.

Valid Sample Entries already defined for ISO and MP4 include MP4AudioSampleEntry and HintSampleEntry. Other Sample Entries shall be according to the following:

- AMR, AMR-WB AMRSampleEntry

- AMR-WB+ AMRWPSampleEntry

- EVS EVSSampleEntry

- IVAS IVASSampleEntry

- H.263 H263SampleEntry

- H.264(AVC) AVCSampleEntry

- H.265(HEVC) HEVCSampleEntry

- Timed text TextSampleEntry

- DIMS DIMSSampleEntry

- CVO timed metadata CVOSampleEntry

- Location timed metadata LocationSampleEntry

- Quality metrics timed metadata QualityMetricsSampleEntry

- Orientation timed metadata OrientationSampleEntry

The format of SampleEntry and its fields are explained as follows:

**SampleEntry ::= MP4AudioSampleEntry |   
 AMRSampleEntry |   
 AMRWPSampleEntry |**

**EVSSampleEntry |  
 IVASSampleEntry |  
 H263SampleEntry |   
 AVCSampleEntry |  
 TextSampleEntry |   
 DIMSSampleEntry |  
 HintSampleEntry |  
 CVOSampleEntry |  
 HEVCSampleEntry |  
 LocationSampleEntry |**

**QualityMetricsSampleEntry |**

**OrientationSampleEntry**

Table 6.1: SampleEntry fields

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Details | Value |
|  |  |  |  |
| MP4AudioSampleEntry |  | Entry type for audio samples defined in the MP4 specification. |  |
| AMRSampleEntry |  | Entry type for AMR and AMR-WB speech samples defined in clause 6.5 of the present document. |  |
| AMRWPSampleEntry |  | Entry type for AMR-WB+ audio samples defined in clause 6.9 of the present document. |  |
| EVSSampleEntry |  | Entry type for EVS samples defined in clause 6.14 of the present document. |  |
| IVASSampleEntry |  | Entry type for IVAS samples defined in clause 6.15 of the present document. |  |
| H263SampleEntry |  | Entry type for H.263 visual samples defined in clause 6.6 of the present document. |  |
| AVCSampleEntry |  | Entry type for H.264 (AVC) visual samples defined in the AVC file format specification in clause 5 of [20]. |  |
| TextSampleEntry |  | Entry type for timed text samples defined in the timed text specification |  |
| DIMSSampleEntry |  | Entry type for DIMS scene description samples defined in the DIMS specification. |  |
| HintSampleEntry |  | Entry type for hint track samples defined in the ISO specification. |  |
| CVOSampleEntry |  | Entry type for CVO timed metadata track as defined in clause 6.11 of the present document |  |
| HEVCSampleEntry |  | Entry type for H.265 (HEVC) visual samples defined in the H.265 (HEVC) file format specification in clause 8 of [20]. |  |
| LocationSampleEntry |  | Entry type for Location timed metadata track as defined in clause 6.12 of the present document |  |
| QualityMetricsSampleEntry |  | Entry type for Quality metrics timed metadata track as defined in clause 4 of [53] |  |
| OrientationSampleEntry |  | Entry type for Orientation timed metadata track as defined in clause 6.13 of the present document |  |

From the Sample Entries in Table 6.1, only the MP4AudioSampleEntry, H263SampleEntry, AMRSampleEntry, AMRWPSampleEntry, EVSSampleEntry, IVASSampleEntry, CVOSampleEntry, LocationSampleEntry and OrientationSampleEntry are taken into consideration here. TextSampleEntry is defined in [4], HintSampleEntry in [7], AVCSampleEntry in clause 5 of [20], HEVCSampleEntry in clause 8 of [20], QualityMetricsSampleEntry in clause 4 of [53] and DIMSSampleEntry in [36].

## 6.14 EVSSampleEntry box

The box type of the EVSSampleEntry Box shall be 'sevs'.

The EVSSampleEntry Box is defined as follows:

**EVSSampleEntry ::= BoxHeader**  
 Reserved\_6  
 Data-reference-index  
 Reserved\_8  
 channelcount  
 Reserved\_2  
 Reserved\_4  
 TimeScale  
 Reserved\_2

Table 6.14: EVSSampleEntry fields

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Type** | **Details** | **Value** |
| **BoxHeader**.Size | Unsigned int(32) |  |  |
| **BoxHeader**.Type | Unsigned int(32) |  | 'sevs' |
| Reserved\_6 | Unsigned int(8) [6] |  | 0 |
| Data-reference-index | Unsigned int(16) | Index to a data reference that to use to retrieve the sample data. Data references are stored in data reference boxes. |  |
| Reserved\_8 | Const unsigned int(32) [2] |  | 0 |
| channelcount | Const unsigned int(16) | Number of mono channels present in this media. |  |
| Reserved\_2 | Const unsigned int(16) |  | 16 |
| Reserved\_4 | Const unsigned int(32) |  | 0 |
| TimeScale | Unsigned int(16) | Sample rate of the media according to the maximum encoded bandwidth, e.g., 32000 for SWB. Set to 48000 if unknown. | One of the values: 8000, 16000, 32000, or 48000. |
| Reserved\_2 | Const unsigned int(16) |  | 0 |

## 6.15 IVASSampleEntry box

The box type of the IVASSampleEntry Box shall be 'sivs'.

The IVASSampleEntry Box is defined as follows:

**IVASSampleEntry ::= BoxHeader**  
 Reserved\_6  
 Data-reference-index  
 Reserved\_8  
 channelcount  
 Reserved\_2  
 Reserved\_4  
 TimeScale  
 Reserved\_2  
 **IVASCodedFormatBox**  
 **IVASVersionBox**

Table 6.15: IVASSampleEntry fields

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Type** | **Details** | **Value** |
| **BoxHeader**.Size | Unsigned int(32) |  |  |
| **BoxHeader**.Type | Unsigned int(32) |  | 'sivs' |
| Reserved\_6 | Unsigned int(8) [6] |  | 0 |
| Data-reference-index | Unsigned int(16) | Index to a data reference that to use to retrieve the sample data. Data references are stored in data reference boxes. |  |
| Reserved\_8 | Const unsigned int(32) [2] |  | 0 |
| channelcount | Const unsigned int(16) | Channel-count of IVAS; in case of Stereo/Binaural audio the expected playback is on 2-channel setups, generally the coded media can be rendered to different speaker setups | May be 2 in case of Stereo/Binaural-, otherwise shall be 0 |
| Reserved\_2 | Const unsigned int(16) |  | 16 |
| Reserved\_4 | Const unsigned int(32) |  | 0 |
| TimeScale | Unsigned int(16) | Sample rate of the media according to the maximum encoded bandwidth, e.g., 32000 for SWB. Set to 48000 if unknown. | One of the values: 16000, 32000, or 48000. |
| Reserved\_2 | Const unsigned int(16) |  | 0 |
| IVASCodedFormatBox |  | Information specific to the VAS decoder to identify the coded format. | 'ivcf' |
| IVASVersionBox |  | Information specific to the VAS decoder to identify the IVAS version used to encode the IVAS samples. | 'iver' |

## 6.16 IVASCodedFormatBox field for IVASSampleEntry box

The IVASCodedFormatBox fields for IVAS shall be as defined in table 6.16. The IVASCodedFormatsBox for the IVASSampleEntry Box shall always be included if the 3GP file contains IVAS media.

Table 6.16: The IVASCodedFormatBox fields for IVASSampleEntry

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Details | Value |
| BoxHeader.Size | Unsigned int(32) |  |  |
| BoxHeader.Type | Unsigned int(32) |  | 'ivcf' |
| IVASFormatInfo | IVASFormatStruc | Structure which describes the coded format of the IVAS media |  |

**BoxHeader Size and Type:** indicate the size and type of the IVAS coded format box. The type must be 'ivcf'.

**IVASFormatInfo:** the structure where the IVAS specific format information resides.

The IVASFormatStruc is defined as follows:

*struct* **IVASFormatStruc**{  
 Unsigned int (16) **codedFormatCategory**

**switch (codedFormatCategory) {  
 case 0 /\*Unknown\*/ :** Unsigned int (16) **reserved;** break; **case 1 /\*Mono\*/ :** Unsigned int (16) **reserved;** break; **case 2: /\*Stereo\*/:** Unsigned int (8) **reserved;** Unsigned int (1) **is\_binaural;** Unsigned int (7) **reserved;** break; **case 3: /\*SBA\*/:** Unsigned int (8) **reserved;** Unsigned int (1) **is\_planar;** Unsigned int (5) **reserved;** Unsigned int (2) **sba\_order;** break; **case 4 /\*MASA\*/ :** Unsigned int (1) reserved;  
 Unsigned int (1) num\_MASA\_transport;  
 Unsigned int (2) src\_format;  
 Unsigned int (12) var\_description; **break;  
 case 5 /\*ISM\*/ :** Unsigned int (8) **reserved;** Unsigned int (6) **reserved;** Unsigned int (4) **num\_objects;** break; **case 6: /\*MC\*/:** Unsigned int (8) **reserved;** Unsigned int (8) **CICP\_idx;  
 case 7: /\*OMASA\*/:** Unsigned int (4) **reserved;** Unsigned int (4) **num\_objects;** Unsigned int (8) **reserved;** break; **case 8: /\*OSBA\*/:**  
 Unsigned int (4) **reserved;** Unsigned int (4) **num\_objects;** Unsigned int (1) **is\_planar;** Unsigned int (5) **reserved;** Unsigned int (2) **sba\_order;** break; **}**}

The definitions of IVASFormatStruc members are as follows:

**codedFormatCategory:** index for the category of the format encoded by the IVAS codec in the samples as indicated in Table 6.17:

Table 6.17: IVAS codedFormatCategory

|  |  |
| --- | --- |
| Index | Indicated Format Category |
| 0 | Unknown; this code point shall only be used if the coded format of the IVAS samples in the file is unknown |
| 1 | Mono using the EVS coding modes of IVAS |
| 2 | Stereo (including Binaural Audio) |
| 3 | Scene-based audio (SBA) |
| 4 | Metadata assisted spatial audio (MASA) |
| 5 | Object-based audio (ISM) |
| 6 | Multi-channel audio (MC) |
| 7 | Combined ISM and MASA (OMASA) |
| 8 | Combined ISM and SBA (OSBA) |

**is\_binaural:** flag that is set to 1 if the coded format is stereo with binaural audio, i.e. meant for presentation over headphones without additional binauralization

**is\_planar:** flag that is set to 1 if the coded format is planar scene-based audio, i.e. consisting of Ambisonics components on the horizontal plane only

**sba\_order:** order of the scene-based audio ranging from 1 to 3.

**num\_objects:** number of audio objects for ISM audio from 1 to 4.

**CICP\_idx:** defined speaker layout from as per Table 3 in [58], which shall be one of the values 6, 12, 14, 16, 19 for 5\_1, 7\_1, 5\_1\_2, 5\_1\_4, 7\_1\_4, respectively.

**num\_MASA\_transport:** number of MASA transport channels, shall be 1 or 2.

**src\_format:** "Source format" from which MASA was created as defined in Annex A.4 of [59].

**var\_description:** "Variable description (12 bits including zero padding)" as defined in Annex A.4 of [59].

**reserved:** unused field reserved for future use; shall be set to 0.

## 6.17 IVASVersionBox field for IVASSampleEntry box

The IVASVersionBox fields for IVAS shall be as defined in table 6.18. The IVASVersionBox for the IVASSampleEntry Box shall always be included if the 3GP file contains IVAS media. Its purpose is to identify the specification version of the IVAS samples in the file to provide a hint to decoders of potential non-interoperability, based on the major version. In IVAS major versions are increased if there are breaking bitstream changes.

Table 6.18: The IVASSpecificBox fields for IVASSampleEntry

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Details | Value |
| BoxHeader.Size | Unsigned int(32) |  |  |
| BoxHeader.Type | Unsigned int(32) |  | 'iver' |
| IVASVersionInfo | IVASVersionStruc | Structure which holds the IVAS version |  |

**BoxHeader Size and Type:** indicate the size and type of the IVAS decoder-specific box. The type must be 'iver'.

**IVASVersionInfo:** the structure where the IVAS version information resides.

The IVASVersionStruc is defined as follows:

*struct* **IVASVersionStruc**{  
 Unsigned int (12) **reserved** Unsigned int (4) **IVAS\_specification**  
 Unsigned int (8) **IVAS\_major\_version** Unsigned int (8) **IVAS\_minor\_version**}

The definitions of IVASVersionStruc members are as follows:

**IVAS\_specification:** index for the specification that was used to enode the sample as indicated in Table 6.19. This field should be set to the specification the samples conform to or to 0 if this is unknown.

Table 6.19: IVAS\_specification

|  |  |
| --- | --- |
| Index | Indicated Format Category |
| 0 | Unknown; |
| 1 | Fixed-point (FX) according to TS 26.251 |
| 2 | Floating-Point (FL) according to TS 26.258 |

**IVAS\_major\_version:** major version of the IVAS specification. This field indicates to decoders which major version has been used for the encoding and shall be set to the major version of the IVAS specification the samples conform to. The information may be used by decoders to hint at potential bitstream non-interoperability between different major versions.

**IVAS\_minor\_version:** minor version of the IVAS specification. This field indicates to decoders which minor version has been used for the encoding. The field should be set to minor version of the IVAS specification the samples conform to.

## A.2.2 Codecs parameter

The codecs parameter is defined in RFC 6381. The ISO file format name space and ISO syntax in clauses 3.3 and 3.4 of RFC 6381 [32] shall be used together with extensions to the ISO syntax specified here.

The syntax in clause 3.4 of RFC 6381 defines the usage of the codecs parameter for files based on the ISO base media file format and specifies that the first element of a parameter value is a sample description entry four-character code. It also includes specific definitions for MPEG audio ('mp4a') where each value in addition to the four-character code includes two elements signalling Object Type Indications and Profile Level Indications (video only). It also includes specific definitions for Advanced Video Coding ('avc1') where each value in addition to the four-character code includes a second element (referred to as 'avcoti' in the formal syntax), which is the hexadecimal representation of the following three bytes in the (subset) sequence parameter set Network Abstraction Layer (NAL) unit specified in [29]: (1) profile\_idc,(2) the byte containing the constraint\_set flags (currently constraint\_set0\_flag through constraint\_set5\_flag, and the reserved\_zero\_2bits), and (3) level\_idc. Note also that reserved\_zero\_2bits is required to be equal to 0 in [29], but other values for it may be specified in the future by ITU-T or ISO/IEC. These definitions apply to the MPEG codecs used by the 3GP file format, such as H.264 (AVC) [29], MPEG-4 AAC [13] and Enhanced aacPlus [23, 24, 25]. Values for other codecs used by the 3GP file format are specified below.

When the first element of a value is 's263', indicating H.263 video [9], the second element is the decimal representation of the profile, e.g., 0 or 3, and the third element is the decimal representation of the level, e.g. 10 or 45.

When the first element of a value is one of the following elements, no other elements are defined for that value:

- 'samr', indicating AMR narrow-band speech [11];

- 'sawb', indicating AMR wide-band speech [12];

- 'sawp', indicating Extended AMR wide-band audio [21];

- 'sevs', indicating EVS speech [55];

- 'sivs', indicating IVAS voice and audio [57];

- 'tx3g', indicating timed text [4];

- '3gvo', indicating CVO as defined in clauses 6.11 and 14.

The following syntax defines all values above in ABNF (RFC 4234 [31]) by extending the definition in clause 3.4 of RFC 6381:

id-iso = iso-gen / iso-mpega / iso-mpegv / iso-amr / iso-amr-wb / iso-amr-wbp / iso-evs / iso-ivas / iso-tt / iso-h263; = iso-gen, iso-mepga, iso-mpegv, iso-avc as defined in RFC 6381

iso-amr = %x73.61.6d.72 ; 'samr'

iso-amr-wb = %x73.61.77.62 ; 'sawb'

iso-amr-wbp = %x73.61.6d.70 ; 'sawp'

iso-evs = %x73.65.76.73 ; 'sevs'

iso-ivas = %x73.65.76.73 ; 'sivs'

iso-tt = %x74.78.33.67 ; 'tx3g'

iso-cvo = %x33.67.76.6f ; '3gvo'

iso-h263 = s263 "." h263-profile "." h263-level

s263 = %x73.32.36.33 ; 's263'

h263-profile = 1\*DIGIT

h263-level = 1\*DIGIT

The elements of the codecs parameter for H.265 (HEVC) are specified as below.

NOTE: The following specification replaces that in clause E.3 of [20].

When the first element of a value is a code indicating a codec from the High Efficiency Video Coding specification (ISO/IEC 23008-2), as documented in clause 8 of [20] (such as 'hev1' or 'hvc1'), the elements following are a series of values from the HEVC decoder configuration record, separated by period characters ("."). In all numeric encodings, leading zeroes may be omitted,

- the general\_profile\_space, encoded as no character (general\_profile\_space == 0), or 'A', 'B', 'C' for general\_profile\_space 1, 2, 3, followed by the general\_profile\_idc encoded as a decimal number;

- the 32 bits of the general\_profile\_compatibility\_flags, but in reverse bit order, i.e. with general\_profile\_compatibility\_flag[ 31 ] as the most significant bit, followed by , general\_profile\_compatibility\_flag[ 30 ], and down to general\_profile\_compatibility\_flag[ 0 ] as the least significant bit, where general\_profile\_compatibility\_flag[ i ] for i in the range of 0 to 31, inclusive, are specified in ISO/IEC 23008-2, encoded in hexadecimal (leading zeroes may be omitted);

- the general\_tier\_flag, encoded as 'L' (general\_tier\_flag==0) or 'H' (general\_tier\_flag==1), followed by the general\_level\_idc, encoded as a decimal number;

- each of the 6 bytes of the constraint flags, starting from the byte containing the general\_progressive\_source\_flag, each encoded as a hexadecimal number, and the encoding of each byte separated by a period; trailing bytes that are zero may be omitted.

Examples:

codecs=hev1.1.6.L93.B0

a progressive, non-packed stream, Main Profile, Main Tier, Level 3.1. (Only one byte of the constraint flags is given here; The value after the second period is 6 instead of 2 because according to ISO/IEC 23008-2 a Main Profile bitstream should also be marked as compatible to the Main 10 Profile).

codecs=hev1.A4.41.H120.B0.23

a (mythical) progressive, non-packed stream in profile space 1, with general\_profile\_idc 4, some compatibility flags set, and in High tier at Level 4 and two bytes of constraint flags supplied.

For Quality metrics metadata track, the usage of the codecs parameter is specified in [53].